The effect of music on auditory perception in cochlear-implant users and normal-hearing listeners

Christina Diechina Fuller

The financial support for printing this thesis is gratefully acknowledged.

Meda Pharma B.V.
Advanced Bionics
Tramedico
Professor dr. Eelco-Huizinga Stichting
Oticon Medical
MED-EL
BCN - Research School of Behavioural and Cognitive Neurosciences
Universitair Medisch Centrum Groningen
Rijksuniversiteit Groningen

ISBN: 978-90-367-9131-1 (book)

Cover: Christina Fuller, Peter van der Sijde
Lay-out: Peter van der Sijde, www.proefschriftgroningen.nl

Printed by: GVO drukkers & vormgevers B.V., Ede

rijksuniversiteit
groningen

The effect of music on auditory perception in cochlear-implant users and normal-hearing listeners

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 26 oktober 2016 om 16.15 uur

door

Christina Diechina Fuller

geboren op 2 juni 1985
te Borger

Promotores

Prof. dr. D. Başkent

Prof. dr. B.F.A.M. van der Laan

Copromotor

Dr. R.H. Free

Beoordelingscommissie

Prof. dr. A. Faulkner

Prof. dr. ir. J.H.M. Frijns

Prof. dr. ir. N.M. Maurits

Paranimfen

Ellen ten Dam

Loes Meijerink

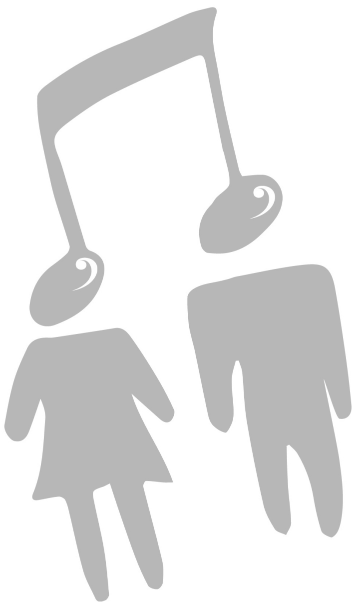
Waar woorden tekort schieten, spreekt de muziek

H.C. Andersen

Contents

Chapter 1

Introduction to the thesis



This thesis focuses on music and speech perception in cochlear implants (CIs). Music and speech are fundamental human communication methods, and are part of every society in the world. Humans begin learning speech and music at a very early age, and both can be extensively trained (especially necessary for music). Both are temporally organized acoustical signals (Asaridou and McQueen 2013). While there are many similarities between speech and music (e.g., pitch and timbre), there are also fundamental differences. Music amongst others targets an emotional response that may differ across listeners, while speech is meant to convey less ambiguous information. Music is structured according to basic elements of pitch, rhythm, timbre, and melody. In terms of lexical meaning ("what is said"), speech is structured in terms of phonemes, syllables, words and sentences. However, speech can also convey indexical cues ("who said it") and prosodic cues ("how it was said") via changes in pitch, timbre and rhythm cues. Speech and music can also be explicitly combined (e.g., sung musical lyrics). Good speech perception is possible using primarily temporal envelope cues (Shannon et al., 1995), but music requires fine-structure cues for harmonic pitch perception (see Fig. 2) (Smith, Delgutte, and Oxenham 2002; Shannon, Fu, and Galvin 2004).

A growing body of research has been directed at understanding neural correlates of music and speech, as well as similarities and differences between speech and music perception. There is also great interest in possible cross-domain effects of long-term musical experience and/or musical training on speech perception (Micheyl et al. 2006; Zatorre 2013; Kraus, Zatorre, and Strait 2014; Herholz and Zatorre 2012; Patel 2014). These studies have been largely conducted with normal hearing (NH) musicians and non-musicians. In NH listeners, the auditory system is intact and capable of perceiving both envelope and fine-structure cues. However, hearing-impaired listeners may not have the same access to fine-structure information, which might limit perception of music and speech where pitch and timbre cues are important. This is especially true for persons who use a CI, an auditory prosthesis to restore hearing to profoundly deaf individuals. Originally designed to convey speech information (i.e., slowly varying spectral and temporal cues), CIs do not effectively transmit fine-structure cues. As a result, music perception is often difficult, as is speech perception, especially in adverse listening conditions (Shannon, Fu, and Galvin 2004). This thesis describes the subjective and behavioral perception of music and pitch-mediated speech in CI users and in NH listeners. We also explored possible advantages of musical experience and training on music and speech perception.

COCHLEAR IMPLANTS

Cochlear implants are auditory prostheses that restore hearing in profoundly deaf individuals by direct stimulation of the auditory neurons using electrodes that are surgically placed within the cochlea. The CI thus provides electrical hearing instead of normal, acoustical hearing. As of 2012, there were 324,200 CI recipients worldwide according to the U.S Food and Drug Administration (National Institute on Deafness and Other Communication Disorder 2011), with approximately 5500 CI recipients in the Netherlands.

CIs were developed to investigate whether electrical stimulation of the auditory nerve could replace acoustical hearing and thereby restore hearing to severely deaf people. The first CIs were single-channel implants, and were implanted into humans in the 1970s (Clark (2003); p15-22). Surprisingly, these patients were capable of some word recognition, but only in combination with lip-reading (Waltzman and Roland 2011). These early studies also showed that the CI improved patients' quality of life (QoL) and the quality of their speech production (Waltzman and Roland 2011). Following these promising early outcomes, CI technology quickly improved, introducing multi-channel stimulation, better electrode designs, and signal processing strategies to reduce noise and improve the transmission of key speech features. As CI technology improved, so did CI outcomes (Blamey et al. 2013). With multi-channel implants, CI users were often capable of audio-only open-set speech recognition. Accordingly, the CI has become accepted worldwide as an effective intervention for post-lingually deafened adults and in many countries, for pre-lingually deafened young children.

SOUND PERCEPTION WITH COCHLEAR IMPLANTS

Despite the success of the CI, there is great variability in patient outcomes and all CI users have difficulty in challenging listening conditions, such as speech understanding in noise, perception of pitch cues in speech, music perception, etc. (Looi, Gfeller, and Driscoll 2012; Gfeller et al. 2008; Fetterman and Domico 2002; Kong et al. 2009; Kong and Carlyon 2010). Some of this variability may be implant-related; some may be patient-related (Blamey et al. 2013; Lazard et al. 2012; Başkent et al. 2016).

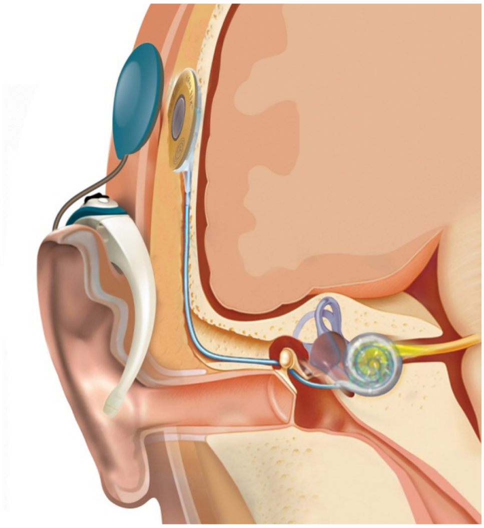


FIGURE 1. A cochlear implant.
http://hearinghealthfoundation.org/lib/sitefiles/images/magazines/CIs_Figure_2_Summer_2012.jpg

Typical CI hardware (see Figure 1) consists of a microphone, a signal processor either body-worn or behind-the ear, a transmitter coil, a receiver coil, and an array of implanted electrodes. The microphone picks up the acoustical signal, which is then optimized and digitized by the signal processor. Signal processing and pre-processing can differ across implant manufacturers. Typically, the acoustic signal is band-pass filtered into frequency analysis bands. The temporal envelope (changes in amplitude over time; see Figure 2) from each band is extracted and used to modulate pulse trains delivered to assigned electrodes. The signal is digitized and transmitted to the receiver coil, which decodes the signal and

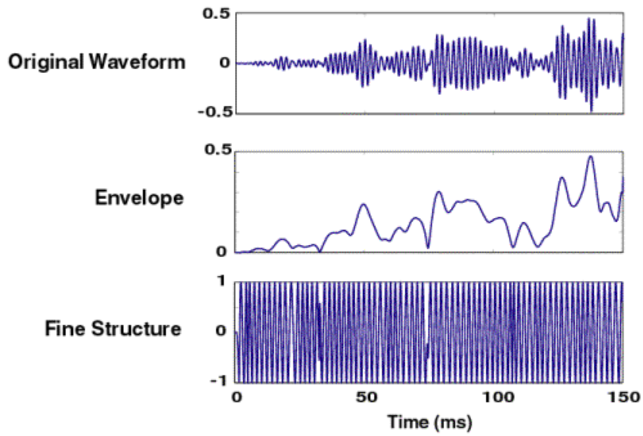


FIGURE 2: The original waveform, the temporal envelope and the fine structure of an acoustical signal.
https://research.meei.harvard.edu/chimera/images/motiva1.gif

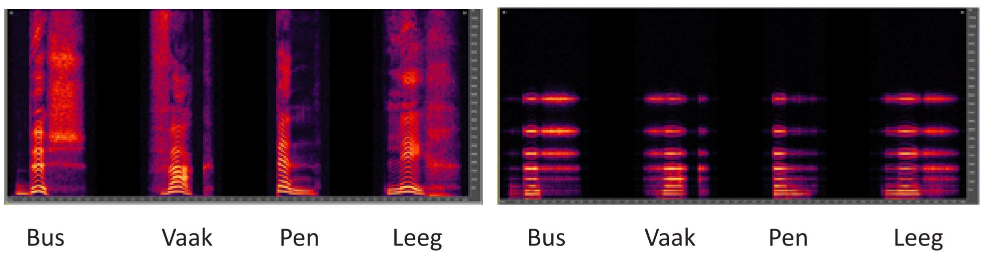


FIGURE 3 (also used in chapter 8) Spectrograms for Dutch monosyllabic words “Bus,” “Vaak,” “Pen,” and “Leeg” (“Bus,” “Often,” “Pen,” and “Empty” in English), shown for unprocessed speech (left panel) or with an 8-channel CI simulation (right panel).

delivers electrical current to the implanted electrodes, thereby directly stimulating the auditory neurons in the spiral ganglia lining the cochlear duct. CI electrode arrays currently have 12 to 22 intra-cochlear electrodes, much fewer than the number of critical bands available for NH listeners to process the wide range of acoustic sounds. Figure 3 shows a spectrogram of unprocessed speech (left panel) and speech processed by an 8-channel CI simulation. The electric dynamic range is also much smaller than in acoustic hearing, making noise problematic for CIs. In typical CI signal processing, spectro-temporal fine structure information is discarded. For most CI users, pitch is perceived via temporal envelope information and changes in the coarse spectral envelope. Due to the limited number of electrodes (12-22) and the interactions among electrodes associated with current spread, complex pitch perception (which requires harmonic frequency components to be resolved) is not presently possible with CIs. Mean frequency discrimination thresholds can be as low as 0.4 semitones for NH listeners, but as high as 5.5 semitones for CI users (Wang, Zhou, and Xu 2011). This poor pitch resolution can greatly limit CI users' melodic pitch perception (Galvin, Fu, and Nogaki 2007; Kong et al. 2004) and perception of vocal emotion or voice gender (Xin, Fu, and Galvin 2007; Gilbers et al. 2015; Gaudrain and Baskent 2015; Fuller et al. 2014c). Beyond implant-related limitations, patient-related factors may further limit perception of the information transmitted by the CI. Duration of deafness, etiology of deafness, patterns of nerve survival, health of auditory neurons, deafness-related changes in cognitive processing may differ across patients, and may explain some of the variability in CI outcomes (Başkent et al. 2016; Blamey et al. 2013)

Thus, implant- and patient-related factors may limit CI users' perception of music and speech. The nature of the listening task and type of stimuli may also play a factor in how well one performs with the CI. In clinical practice, CI performance is only assessed for speech perception, and is often measured using identification of simple sentences and/or monosyllabic words in quiet. CI performance can deteriorate in the presence of steady noise, and further worsens in competing speech or fluctuating maskers (Friesen et al. 2001; Nelson and Jin 2004; Stickney et al. 2004; Nogaki, Fu, and Galvin 2007; Fu and Nogaki 2005). Voice gender identification depends strongly on perception of pitch cues, and can thus be difficult for CI users (Fuller et al. 2014c; Xin, Fu, and Galvin 2007; Wilkinson et al. 2013; Fu, Chinchilla, and Galvin 2004). Vocal emotion identification similarly depends strongly on voice pitch cues and is therefore difficult for many CI users (Xin, Fu, and Galvin 2007; Gilbers et al. 2015). Melodic pitch perception has been shown in many studies to be difficult for CI users (e.g., Gfeller et al. 2007; Galvin et al. 2012; Galvin, Fu, and Nogaki 2007). Thus, different listening tasks and stimuli may elicit further differences among CI users, and better define perceptual limits for "real-life" listening conditions.

MUSIC AND CIs

Music is a fundamental, powerful, and often pleasurable form of human communication (Koelsch et al. 2006; Zatorre and Salimpoor 2013; Salimpoor et al. 2009). Moreover, music is considered to be the second most important acoustical signal after speech by CI users (Boucher and Bryden 1997; Drennan and Rubinstein 2008; Salimpoor et al. 2009; Patel 2014). In many ways, music is a more complicated signal than speech. In terms of perception, fine-structure cues are considered as more important for music and envelope cues more important for speech (Smith, Delgutte, and Oxenham 2002).

There can be great variability among CI users' music enjoyment (subjective measures) and music perception (behavioral measures), but an association between music enjoyment and perception is not a given. Especially music enjoyment seems to be affected by more factors than just perception quality. For example, pre-lingually deafened and early implanted CI children greatly enjoy music, even if their melodic pitch perception is poor (Trehub, Vongpaisal, and Nakata 2009). Post-lingually deafened adult CI users often rate the way music sounds with their CI as poorer than previously experienced with NH (Trehub, Vongpaisal, and Nakata 2009; Gfeller et al. 2000b; Lassaletta et al. 2008). While music perception may not improve, music enjoyment may improve in time (especially with training) and may greatly benefit CI users. For patients with Parkinson's disease and/or dementia, music therapy and training have been shown to improve QoL (Hilliard 2003; Walworth et al. 2008).

Given the potential benefits of music listening and training, efforts to improve music enjoyment could be beneficial in the rehabilitation/or training of CI users. Theoretically, music enjoyment may be increased by improving music perception, as better music perception may provide more enjoyment. Perception of musical pitch, melody and timbre has been shown to be poorer in CI users than in NH listeners (Drennan et al. 2015; Drennan and Rubinstein 2008; McDermott 2004; Limb and Roy 2014). However, musical rhythm is perceived with almost the same accuracy in CI users as in NH listeners (Gfeller, et al. 2007, Kong, et al. 2004). While 4 spectral channels can provide good understanding of speech in quiet, more than 48 spectral channels are needed for melody recognition, and many more channels for good sound quality (Shannon, Fu, and Galvin 2004). Music training may help to compensate for some of the coarse and/or distorted cues provided by the CI. It remains a great challenge in research and development of CIs to sufficiently increase the quality of the signal transmitted to support good pitch perception, which is needed for good music perception.

MUSICIAN EFFECT

Music is a potent acoustical stimulus that can communicate emotions and have positive effects on QoL in NH people. Long-term musical training can also enhance the perception of some acoustical signals. A number of perceptual advantages in musicians have been observed, such as enhanced decoding of emotion in a vocal sound (Wong et al. 2007; Musacchia, Strait, and Kraus 2008; Strait et al. 2009; Besson, Chobert, and Marie 2011), better perception of voicing cues in speech and pitch cues in speech and music (Schon, Magne, and Besson 2004; Thompson, Schellenberg, and Husain 2004; Chartrand and Belin 2006), and better speech understanding in noise (Parbery-Clark et al. 2009; Kraus and

Chandrasekaran 2010). In these studies, these advantages were mostly attributed to long-term musical training. This 'musician effect' is especially interesting as it implies a possible cross-domain transfer of learning from music training to speech perception in NH.

There are different theories regarding the source of the musician effect. One theory is that musicians have better overall pitch perception, suggesting a musician advantage at the lower levels of the auditory system that makes it easier to differentiate the acoustic cues in complex signals (Micheyl et al. 2006; Besson et al. 2007; Oxenham 2008; Deguchi et al. 2012). Another theory is that musicians have a better higher-level processing (e.g., better use of auditory attention, better short- and/or long-term auditory memory) that leads to improved use of cognitive mechanisms for auditory perception and discrimination (Bialystok and Depape 2009; Besson, Chobert, and Marie 2011; Moreno et al. 2011; Barrett et al. 2013).

The areas of auditory perception where musicians seem to show an advantage over non-musicians are precisely the areas in which CI users experience difficulties, and most involve pitch perception. As discussed above, there is great variability in CI users' music enjoyment and perception, as well as great variability in challenging speech perception tasks (e.g., vocal emotion identification, speech-on-speech masking, and voice gender identification). For post-lingually deafened CI users, having music experience before implantation may have partially contributed to this variability. For all CI users music training after implantation may help to improve music and speech perception.

AUDITORY TRAINING IN CI USERS

Most current CI rehabilitation programs are focused on speech perception and production. For post-lingually deafened adults, much of the adaptation to the CI occurs during the first 6-12 months of use, peaking approximately 3.5 years after implantation (Blamey et al. 2013; Rouger et al. 2007). Most CI centers offer a three-month rehabilitation program, after which CI users must adapt via daily exposure to different sounds. Speech training has been shown to be effective in improving CI users' speech perception in quiet and in noise, even after many years of previous experience with their device (Fu, Nogaki, and Galvin III 2005; Stacey et al. 2010; Oba, Fu, and Galvin 2011; Fu and Galvin 2008; Benard and Baskent 2013).

The benefits of music training in CI users have received less attention, as speech perception has long been the main outcome for cochlear implantation. Music training in adult and pediatric CI users has been shown to improve melodic contour identification, timbre recognition, and complex melody recognition (Fu et al. 2015; Galvin et al. 2012; Galvin, Fu, and Nogaki 2007; Yucel, Sennaroglu, and Belgin 2009; Gfeller et al. 2002b; Gfeller et al. 2000b). These previous studies have focused on within-domain (i.e., music perception only) learning and neural plasticity. In this thesis, we explored the possibility of a cross-domain transfer of learning: that training with music could improve both music and speech perception. A previous pilot study by Patel (2014) with two CI users showed a small effect of music training on perception of speech in noise and prosody in words, suggesting some possibility of cross-domain learning with music training.

OUTLINE OF THE THESIS

In this thesis, the perception of music and the effect of music training on auditory perception in CI users and NH listeners were investigated to answer the following research questions:

1. *Can long-term music experience lead to better perception of the degraded signals provided by the CI? If so, this would suggest that music training may benefit perception of degraded signals provided by CIs.*
2. *Can training after implantation benefit CI users' speech and music perception?*
3. *Which music training methods are most effective in CI users?*

The thesis is composed of three parts to systematically explore the research questions listed above:

1. *Assessment of perception of music and pitch-mediated speech stimuli in CI users*
2. *The musician effect in NH listeners and CI users*
3. *The effect of musical training and music therapy in CI users*

1. *Assessment of the perception of music and pitch-mediated speech stimuli in CI users*

First, we assessed the difficulties that CI users experience in enjoying music and perceiving music and pitch-mediated speech, using subjective and behavioral measures. We also investigated potential links between these difficulties to general speech perception and QoL. This first part of the thesis consists of four studies. The first two investigated self-reported music perception and enjoyment in two groups of CI users: 1) early-deafened, late-implanted CI users, and 2) post-lingually deafened CI users. We also examined how self-reported music perception and enjoyment relates to speech perception and QoL. The third and fourth studies further investigated two separate elements; voice gender and vocal emotion perception by NH and CI listeners. In the voice gender categorization study, we manipulated the voice characteristics to gradually change from a female to a male talker, and the task was to identify the speaker's gender. In the vocal emotion identification study, we used a nonsense word produced in four emotions (anger, sadness, joy, and relief), and the task was to identify the correct emotion.

2. *The musician effect in NH listeners and CI users*

Here, we investigated whether active musical training might contribute to a better perception of speech, pitch-mediated speech, and music in CI users and NH musicians and non-musicians. Musicians were used as a model of long-term music training. This part of the thesis consists of three studies. The first study investigated the effect of CI users' music experience and training before implantation on speech perception after implantation. The second and third studies aimed to investigate whether the musician effect persists under conditions of reduced spectro-temporal resolution as experienced by CI users. The second study measured perception of word and sentence intelligibility in quiet and in noise, vocal emotion identification, voice gender categorization, and melodic contour identification in NH musicians and non-musicians. The third study investigated the musician effect in NH subjects' voice gender categorization while listening to unprocessed acoustical stimuli and CI simulations.

3. *The effect of music therapy and music training in CI users*

Here, we directly investigated the effects of music therapy and music training on a group of CI users. This was a prospective, feasibility training study in post-lingually deafened, adult CI users. Outcomes for three different training methods (individualized music training, group music therapy, and non-musical training) were compared in terms of speech intelligibility, music perception, perception of pitch-mediated speech, and QoL. CI users were tested before and after six weeks of training.

REFERENCES

- Asaridou, S. S., and J. M. McQueen. 2013. Speech and music shape the listening brain: Evidence for shared domain-general mechanisms.
- Barrett, K. C., R. Ashley, D. L. Strait, and N. Kraus. 2013. Art and science: How musical training shapes the brain. *Frontiers in Psychology* 4 (Oct 16): 713.
- Başkent, D., E. Gaudrain, T.N. Tamati, and A. Wagner. 2016. Perception and psychoacoustics of speech in cochlear implant users. . In *Scientific foundations of audiology.*, eds. A. T. Cacace, E. de Kleine, A. Holt and P. van Dijk.
- Benard, M. R., and D. Başkent. 2013. Perceptual learning of interrupted speech. *PloS One* 8 (3): e58149.
- Besson, M., J. Chobert, and C. Marie. 2011. Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology* 2 (May 12): 94.
- Besson, M., D. Schon, S. Moreno, A. Santos, and C. Magne. 2007. Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience* 25 (3-4): 399-410.
- Bialystok, E., and A. M. Depape. 2009. Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology* 35 (2) (Apr): 565-74.
- Blamey, P., F. Artieres, D. Başkent, F. Bergeron, A. Beynon, E. Burke, N. Dillier, et al. 2013. Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiology & Neuro-Otology* 18 (1): 36-47.
- Boucher, R., and M. P. Bryden. 1997. Laterality effects in the processing of melody and timbre. *Neuropsychologia* 35 (11) (Nov): 1467-73.
- Chartrand, J. P., and P. Belin. 2006. Superior voice timbre processing in musicians. *Neuroscience Letters* 405 (3) (Sep 25): 164-7.
- Clark, G. 2003. *Cochlear implants: Fundamentals and applications*. AIP series in modern acoustics and signal processing. New York: AIP Press/Springer.
- Deguchi, C., M. Boureux, M. Sarlo, M. Besson, M. Grassi, D. Schön, and L. Colombo. 2012. Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research* 1455 : 75-89.
- Dorman, M. F., P. C. Loizou, and D. Rainey. 1997. Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *The Journal of the Acoustical Society of America* 102 (5): 2993-6.
- Drennan, W. R., J. J. Oleson, K. Gfeller, J. Crosson, V. D. Driscoll, J. H. Won, E. S. Anderson, and J. T. Rubinstein. 2015. Clinical evaluation of music perception, appraisal and experience in cochlear implant users. *International Journal of Audiology* 54 (2) (Feb): 114-23.
- Drennan, W. R., and J. T. Rubinstein. 2008. Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of Rehabilitation Research and Development* 45 (5): 779-89.
- Fetterman, B. L., and E. H. Domico. 2002. Speech recognition in background noise of cochlear implant patients. *Otolaryngology-Head and Neck Surgery* 126 (3) (Mar): 257-63.
- Friesen, L. M., R. V. Shannon, D. Başkent, and X. Wang. 2001. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America* 110 (2) (Aug): 1150-63.
- Fu, Q. J., S. Chinchilla, and John J. Galvin. 2004. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology* 5 (3): 253-60.
- Fu, Q. J., J. J. Galvin, X. Wang, and J. L. Wu. 2015. Benefits of music training in mandarin-speaking pediatric cochlear implant users. *Journal of Speech, Language, and Hearing Research* 58 (1): 163-9.
- Fu, Q. J., and G. Nogaki. 2005. Noise susceptibility of cochlear implant users: The role of spectral resolution

and smearing. *Journal of the Association for Research in Otolaryngology* 6 (1): 19-27.

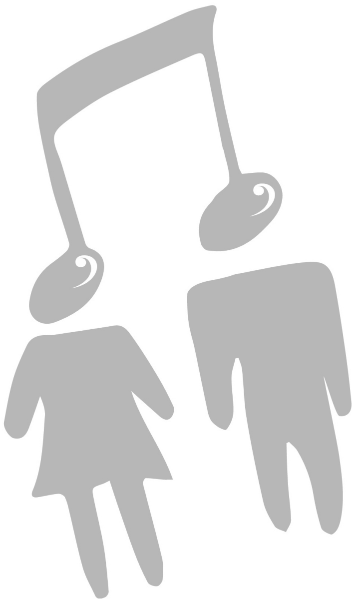
- Fu, Q. J., G. Nogaki, and J. J. Galvin III. 2005. Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation. *Journal of the Association for Research in Otolaryngology* 6 (2): 180-9.
- Fu, Q. J., and J. J. Galvin 3rd. 2008. Maximizing cochlear implant patients' performance with advanced speech training procedures. *Hearing Research* 242 (1-2) (Aug): 198-208.
- Fuller, C. D., E. Gaudrain, J. N. Clarke, J. J. Galvin, Q. J. Fu, R. H. Free, and D. Başkent. 2014. Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology : JARO* 15 (6) (Dec): 1037-48.
- Galvin, J. J., E. Eskridge, S. Oba, and Q. J. Fu. 2012. Melodic contour identification training in cochlear implant users with and without a competing instrument. Paper presented at Seminars in hearing, .
- Galvin, J. J.,3rd, Q. J. Fu, and G. Nogaki. 2007. Melodic contour identification by cochlear implant listeners. *Ear and Hearing* 28 (3) (Jun): 302-19.
- Audrain, E., and D. Başkent. 2015. Factors limiting vocal-tract length discrimination in cochlear implant simulations. *The Journal of the Acoustical Society of America* 137 (3) (Mar): 1298-308.
- Gfeller, K., A. Christ, J. F. Knutson, S. Witt, K. T. Murray, and R. S. Tyler. 2000. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *Journal of the American Academy of Audiology* 11 (7) (Jul-Aug): 390-406.
- Gfeller, K., J. Oleson, J. F. Knutson, P. Breheny, V. Driscoll, and C. Olszewski. 2008. Multivariate predictors of music perception and appraisal by adult cochlear implant users. *Journal of the American Academy of Audiology* 19 (2) (Feb): 120-34.
- Gfeller, K., C. Turner, J. Oleson, X. Zhang, B. Gantz, R. Froman, and C. Olszewski. 2007. Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear and Hearing* 28 (3) (Jun): 412-23.
- Gfeller, K., S. Witt, M. Adamek, M. Mehr, J. Rogers, J. Stordahl, and S. Ringgenberg. 2002. Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *Journal of the American Academy of Audiology* 13 (3) (Mar): 132-45.
- Gilbers, S., C. Fuller, D. Gilbers, M. Broersma, M. Goudbeek, R. Free, and D. Başkent. 2015. Normal-hearing listeners' and cochlear implant users' perception of pitch cues in emotional speech. *I-Perception* 6 (5): 0301006615599139.
- erholz, S. C., and R. J. Zatorre. 2012. Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron* 76 (3) (Nov 8): 486-502.
- Hilliard, R. E. 2003. The effects of music therapy on the quality and length of life of people diagnosed with terminal cancer. *Journal of Music Therapy* 40 (2) (Summer): 113-37.
- Koelsch, S., T. Fritz, K. Müller, and A. D. Friederici. 2006. Investigating emotion with music: An fMRI study. *Human Brain Mapping* 27 (3): 239-50.
- Kong, Y. Y., and R. P. Carlyon. 2010. Temporal pitch perception at high rates in cochlear implants. *The Journal of the Acoustical Society of America* 127 (5) (May): 3114-23.
- Kong, Y. Y., R. Cruz, J. A. Jones, and F. G. Zeng. 2004. Music perception with temporal cues in acoustic and electric hearing. *Ear and Hearing* 25 (2) (Apr): 173-85.
- Kong, Y. Y., J. M. Deeks, P. R. Axon, and R. P. Carlyon. 2009. Limits of temporal pitch in cochlear implants. *The Journal of the Acoustical Society of America* 125 (3) (Mar): 1649-57.
- Kraus, N., and B. Chandrasekaran. 2010. Music training for the development of auditory skills. *Nature Reviews. Neuroscience* 11 (8) (Aug): 599-605.
- Kraus, N., R. J. Zatorre, and D. L. Strait. 2014. Editors' introduction to hearing research special issue: Music: A window into the hearing brain. *Hearing Research* 308 (Feb): 1.

- Lassaletta, L., A. Castro, M. Bastarrica, R. Perez-Mora, B. Herran, L. Sanz, M. J. de Sarria, and J. Gavilan. 2008. Musical perception and enjoyment in post-lingual patients with cochlear implants. *Acta Otorrinolaringologica Espanola* 59 (5) (May): 228-34.
- Lazard, D. S., C. Vincent, F. Venail, P. Van de Heyning, E. Truy, O. Sterkers, P. H. Skarzynski, et al. 2012. Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: A new conceptual model over time. *PloS One* 7 (11): e48739.
- Limb, C. J., and A. T. Roy. 2014. Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing Research* 308 (Feb): 13-26.
- Looi, V., K. Gfeller, and V. Driscoll. 2012. Music appreciation and training for cochlear implant recipients: A review. *Seminars in Hearing* 33 (4) (Nov 1): 307-34.
- McDermott, H. J. 2004. Music perception with cochlear implants: A review. *Trends in Amplification* 8 (2): 49-82.
- Micheyl, C., K. Delhommeau, X. Perrot, and A. J. Oxenham. 2006. Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research* 219 (1-2) (Sep): 36-47.
- Moreno, S., E. Bialystok, R. Barac, E. G. Schellenberg, N. J. Cepeda, and T. Chau. 2011. Short-term music training enhances verbal intelligence and executive function. *Psychological Science* 22 (11) (Nov): 1425-33.
- Musacchia, G., D. Strait, and N. Kraus. 2008. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research* 241 (1-2) (Jul): 34-42.
- National Institute on Deafness and Other Communication Disorder. 2011. “Cochlear implants”. *NIH Publication no. 11-4798 (2013-11-01).*
- Nelson, P. B., and S. H. Jin. 2004. Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America* 115 (5): 2286-94.
- Nogaki, G., Q. J. Fu, and J. J. Galvin 3rd. 2007. Effect of training rate on recognition of spectrally shifted speech. *Ear and Hearing* 28 (2) (Apr): 132-40.
- Oba, S. I., Q. J. Fu, and J. J. Galvin 3rd. 2011. Digit training in noise can improve cochlear implant users’ speech understanding in noise. *Ear and Hearing* 32 (5) (Sep-Oct): 573-81.
- Oxenham, A. J. 2008. Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification* 12 (4) (Dec): 316-31.
- Parbery-Clark, A., E. Skoe, C. Lam, and N. Kraus. 2009. Musician enhancement for speech-in-noise. *Ear and Hearing* 30 (6) (Dec): 653-61.
- Patel, A. D. 2014. Can nonlinguistic musical training change the way the brain processes speech? the expanded OPERA hypothesis. *Hearing Research* 308 (Feb): 98-108.
- Rouger, J., S. Lagleyre, B. Fraysse, S. Deneve, O. Deguine, and P. Barone. 2007. Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proceedings of the National Academy of Sciences of the United States of America* 104 (17) (Apr 24): 7295-300.
- Salimpoor, V. N., M. Benovoy, G. Longo, J. R. Cooperstock, and R. J. Zatorre. 2009. The rewarding aspects of music listening are related to degree of emotional arousal. *PloS One* 4 (10) (Oct 16): e7487.
- Schon, D., C. Magne, and M. Besson. 2004. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* 41 (3) (May): 341-9.
- Shannon, R. V., Q. J. Fu, and J. Galvin 3rd. 2004. The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Oto-Laryngologica.Supplementum* (552) (552) (May): 50-4.
- Shannon, R. V., F. G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid. 1995. Speech recognition with primarily temporal cues. *Science (New York, N.Y.)* 270 (5234) (Oct 13): 303-4.
- Smith, Zachary M., Bertrand Delgutte, and Andrew J. Oxenham. 2002. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416 (6876): 87-90.

- Stacey, P. C., C. H. Raine, G. M. O'Donoghue, L. Tapper, T. Twomey, and A. Q. Summerfield. 2010. Effectiveness of computer-based auditory training for adult users of cochlear implants. *International Journal of Audiology* 49 (5) (May): 347-56.
- Stickney, G. S., F. G. Zeng, R. Litovsky, and P. Assmann. 2004. Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America* 116 (2) (Aug): 1081-91.
- Strait, D. L., N. Kraus, E. Skoe, and R. Ashley. 2009. Musical experience promotes subcortical efficiency in processing emotional vocal sounds. *Annals of the New York Academy of Sciences* 1169 (Jul): 209-13.
- Thompson, W. F., E. G. Schellenberg, and G. Husain. 2004. Decoding speech prosody: Do music lessons help? *Emotion (Washington, D.C.)* 4 (1) (Mar): 46-64.
- Trehub, S. E., T. Vongpaisal, and T. Nakata. 2009. Music in the lives of deaf children with cochlear implants. *Annals of the New York Academy of Sciences* 1169 (Jul): 534-42.
- Waltzman, Susan B., and J. Thomas Roland. 2011. *Cochlear implants.* Thieme, New York, 2nd edition.
- Walworth, D., C. S. Rumana, J. Nguyen, and J. Jarred. 2008. Effects of live music therapy sessions on quality of life indicators, medications administered and hospital length of stay for patients undergoing elective surgical procedures for brain. *Journal of Music Therapy* 45 (3) (Fall): 349-59.
- Wang, W., N. Zhou, and L. Xu. 2011. Musical pitch and lexical tone perception with cochlear implants. *International Journal of Audiology* 50 (4): 270-8.
- Wilkinson, E. P., O. Abdel-Hamid, J. J. Galvin, H. Jiang, and Q. J. Fu. 2013. Voice conversion in cochlear implantation. *The Laryngoscope* 123 (S3): S29-43.
- Wong, P. C., E. Skoe, N. M. Russo, T. Dees, and N. Kraus. 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience* 10 (4) (Apr): 420-2.
- Xin, Luo, Q. J. Fu, and J. J. Galvin 3rd. 2007. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification* 11 (4) (Dec): 301-15.
- Yucel, Esra, Gonca Sennaroglu, and Erol Belgin. 2009. The family oriented musical training for children with cochlear implants: Speech and musical perception results of two year follow-up. *International Journal of Pediatric Otorhinolaryngology* 73 (7): 1043-52.
- Zatorre, R. J. 2013. Predispositions and plasticity in music and speech learning: Neural correlates and implications. *Science (New York, N.Y.)* 342 (6158) (Nov 1): 585-9.
- Zatorre, R. J., and V. N. Salimpoor. 2013. From perception to pleasure: Music and its neural substrates. *Proceedings of the National Academy of Sciences of the United States of America* 110 Suppl 2 (Jun 18): 10430-7.

Part one:

Assessment of the perception of music and pitch-mediated speech stimuli in cochlear-implant users

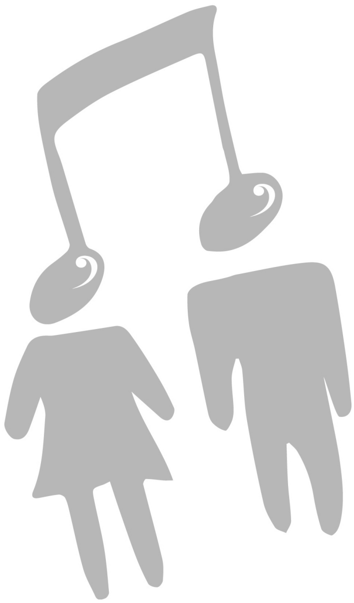


Chapter 2

Music and quality of life in early-deafened late-implanted adult cochlear implant users

Christina Fuller [1,2], Lisa Mallinckrodt [1,3], Bert Maat [1,2], Deniz Başkent [1,2], and Rolien Free [1,2]

Otology & Neurotology 34:1041-1047 2013



[1] Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[2] Graduate School of Medical Sciences (Research School of Behavioural and Cognitive Neurosciences), University of Groningen, The Netherlands.

[3] Department of Surgery, Scheper Ziekenhuis, Emmen, The Netherlands

ABSTRACT

Hypothesis and Background: The early-deafened, late-implanted (EDLI) CI users constitute a relatively new and understudied clinical population. To contribute to a better understanding of the implantation outcome, this study evaluated this population for self-reported enjoyment and perception of music. Additionally, correlations of these measures with the self-reported quality of life and everyday hearing ability, and a behaviorally measured word recognition test were explored.

Materials and Methods: EDLI CI users from the Northern-Netherlands were sent four questionnaires: 1) Dutch Musical Background Questionnaire (enjoyment and perception of music); 2) Nijmegen Cochlear Implant Questionnaire (quality of life); 3) Cochlear Implant Functioning Index (auditory-related functioning); 4) Speech, Spatial and Qualities of Hearing Scale (hearing ability). Complementary, behavioral word recognition in quiet tests (phoneme score) were completed.

Results: Twelve out of 20 (60%) participants reported music to sound pleasant. In general, the self-perceived quality of music was scored positively. No correlations were observed between enjoyment and perception of music, quality of life, hearing ability and word recognition.

Conclusion: The results indicate that, differently than post-lingually deafened, EDLI CI users enjoy music and rate the quality of music positively. Potential explanations for the absence of correlations between the music measures and the other outcomes could be that other factors, such as speech perception, contribute more to quality of life of EDLI CI users or that this group simply lacks previous exposure to music with acoustical hearing. Overall, these positive findings may give extra support for implant candidacy of early-deafened individuals, but further studies should be conducted.

Keywords: Cochlear Implant, Early Deafened, Late Implanted Adults, Music Perception, Music enjoyment, Quality of Life

INTRODUCTION

Cochlear implants (CIs) restore hearing in severely deafened adults and children. Nowadays, the perception of speech in quiet is fairly good in post-lingually deafened CI-users. Nevertheless, the perception of music is still inadequate and dissatisfactory in this population (Boucher and Bryden 1997; Drennan and Rubinstein 2008; Gfeller et al. 2000b; Fuller et al. 2012; Fuller et al. under revision; Looi and She 2010; Migirov, Kronenberg, and Henkin 2009). A potential explanation for this dissatisfaction could be that the perception of music (or the processing of its four basic elements; pitch, rhythm, melody and timbre) is less accurate and more variable in CI users compared to normal hearing (NH) listeners (Drennan and Rubinstein 2008; McDermott 2004; Wang et al. 2012). Interestingly, early-deafened, early-implanted (EDEI) young CI users report higher enjoyment of listening to music than post-lingually deafened adult CI users, even though the behaviorally measured ability of music perception in EDEI has been observed to be worse (Mitani et al. 2007; Jung et al. 2012; Vongpaisal, Trehub, and Schellenberg 2009). Based on the observations with EDEI population, early-deafened, late implanted (EDLI; late implantation defined as after the age of sixteen in the present study) CI users may also have a different appreciation of music than the post-lingually implanted group.

Despite a delay between the onset of deafness and the implantation, which has negative consequences for the speech perception outcome in general (Lazard et al. 2012; Blamey et al. 2013), and a potential deficit in language skills due to the onset of deafness in early childhood, a subgroup of EDLI CI users have been observed to benefit from implantation regarding speech perception and quality of life (QoL) (Mallinckrodt et al. ; Klop et al. 2007; Niparko et al. 2010; Houston and Miyamoto 2010; Santarelli et al. 2008; De Raeve 2010; Most, Shrem, and Duvdevani 2010; Yoshida et al. 2008; Schramm, Fitzpatrick, and Séguin 2002). However, the benefit for enjoyment and perception of music in this group is mostly unknown. To the best of our knowledge, only two studies have examined the self-reported perception of music in EDLI, and both presented some limitations (Migirov, Kronenberg, and Henkin 2009; Eisenberg 1982). While both reported that EDLI CI users enjoy listening to music using their implant, in the study by Migirov et al. (2009) with nine pre-lingually deafened CI users the age at implantation was unknown and the study by Eisenberg with twelve CI users was published in the early days of the CIs in 1982. As the CI technology, surgical techniques, rehabilitation methods, as well as CI outcome, have substantially changed since then, an updated and a more comprehensive evaluation of this group is needed.

Music is a pleasurable stimulus that can affect emotional states, to the degree that music therapies can positively influence QoL (Salimpoor et al. 2011; Hilliard 2003). Therefore, any improvement to the perception of music could presumably have similar positive effects for CI users. Hence, assessing the perception and enjoyment of music in the understudied group of EDLI CI users could give additional insight in the debate on whether or not implantation of early deafened adults or adolescents would still be beneficial at a later age. Moreover, the perception and enjoyment of music could influence other outcome factors of implantation, such as QoL, everyday hearing ability and speech perception (Fuller et al. 2012; Fuller et al. under revision; Lassaletta et al. 2007; Lassaletta et al. 2008). In the present study we have explored the self-reported enjoyment and perception of music in EDLI CI users more extensively and systematically than the two previous studies, by collecting extensive data on demographics and patient history, and careful selection of the participants accordingly.

Complementary correlational analyses between the self-reported perception and enjoyment of music, the health-related quality of life (HRQoL), everyday hearing ability and a behavioral word recognition measure were explored. We first hypothesized that EDLI CI users would enjoy listening to music, similar to EDEI CI users, but unlike the post-lingually deafened CI users. Second, we also hypothesized that higher enjoyment and better perception of music would be correlated with higher QoL, better everyday hearing ability and word recognition, based on the findings with post-lingually deafened CI users (Fuller et al. 2012; Fuller et al. under revision; Lassaletta et al. 2007).

MATERIALS AND METHODS

Study population

The inclusion criteria for participation, based on (Mallinckrodt et al. ; Klop et al. 2007; Goorhuis-Brouwer and Schaerlaekens 2000), were: severe hearing loss at least since preschool (onset six years of age or earlier), implanted at 16 years of age or later, and more than one year of CI-experience. The criterion sixteen years or later was picked to assure a period of auditory deprivation in the EDLI. Thirty-seven qualifying EDLI CI users, all patients of our clinic and a subgroup of the participants previously described by Mallinckrodt et al. (2012), were sent four questionnaires. Twenty-seven (73%) replies were received. Five CI users were excluded after their responses revealed that they did not strictly meet the inclusion criterion for severe hearing loss onset at the age of six or earlier. The demographics of the 22 study participants are shown in Table I. The levels of education refer to the highest completed educational level: low refers to elementary school only; middle refers to middle school or higher; high refers to at least a bachelor's degree.

The study was approved by the Medical Ethical Committee of the University Medical Center Groningen. Participants were given detailed information about the study and written informed consent was obtained. Participation was entirely voluntary and no financial reimbursement was provided.

Dutch Musical Background Questionnaire

The Dutch Musical Background Questionnaire (DMBQ) is a translated and edited version of the Iowa Musical Background Questionnaire (Gfeller et al. 2000b). The questionnaire

TABLE I: Demographics of all study participants. N refers to the number of participants in this and following tables and figures.

All participants N = 22			
	Number	Mean (Standard deviation)	Range
Gender			
Male	6		
Female	16		
Age (y)		47.4 ± 15.0	19-68
Age at onset of severe hearing loss (y)*		0.7±1.3	0-4
Age at fitting of first hearing aid (y)*		2.5±1.5	0-8
Age at implantation (y)		41.2±14.3	17-63
Level of education			
Lower	3		
Middle	12		
Higher	7		
Deaf school attendance			
Sign language school	9		
Aural communication/sign language school	8		
Aural/oral school	5		
Duration of CI use (y)*		5.7±3.3	1-10
CI use per day (h)*		14.2 ± 4.2	6-24
Implant type			
CI24R CA[a]	4		
CI24R k[a]	1		
CI24RE CA[a]	5		
CI24R CS[a]	7		
HiRes90K Helix[b]	5		
Speech processor type (no.)			
Esprit3G[a]	5		
Freedom[a]	7		
Nucleus 5[a]	5		
Harmony[b]	5		

a Cochlear Corp., Englewood, Australia device. ACE speech strategy. b Advanced Bionics Corp., California, USA device. HiRes speech strategy. y= years. h=hours.
* Based on patient reports

was translated into Dutch by a professional translator with assistance from the first author, and was further revised by an audiologist, an Ear-, Nose- and Throat surgeon, audiology scientists and a psychologist. The DMBQ has three parts that measure: satisfaction with listening to music, self-perceived quality of music and self-reported perception of the elements of music.

Satisfaction with listening to music

The satisfaction with listening to music was determined via a three option single question: *little or no satisfaction with listening to music*; *the sound of music is okay or improving over time*; *music sounds pleasant*. The satisfaction was accordingly scored on a 0 (no satisfaction) to 2 (most satisfaction) scale by 20 (out of 22) CI users. Note that not every respondent filled all questions of all questionnaires. Therefore the number of participants is specified in all results and figures.

Self-perceived quality of music

The self-perceived quality of music is an indication of how music sounds under the best conditions with a CI. Twenty-two respondents scored seven visual analog scales (VASs) with fourteen opposite adjective descriptors (*unpleasant-pleasant*, *mechanical-natural, fuzzy-clear, does not sound like music-sounds like music, complex-simple, difficult to follow-easy to follow, dislike very much-like very much*). The scales ranged from 0 (negative quality) to 100 (positive quality). An average across the seven scales was taken to quantify the self-perceived quality of music.

Self-reported perception of the elements of music

Participants reported their ability to perceive the elements of music: rhythm, melody and timbre, and to differentiate between vocalists and lyrics. The specific questions were:

1. Can you hear the difference between singing and speaking?
2. Are you able to differentiate between a male and a female vocalist?
3. Are you able to follow the rhythm of a music piece?
4. Are you able to recognize the melody of a music piece?
5. Are you able to differentiate the instruments in a piece of music?
6. Can you follow the lyrics of a song?

The six questions were scored on a scale from 1 (never) to 7 (always). The scores 1 to 3 were classified as a 'negative' ability, 4 as a 'neutral' ability and 5 to 7 as a 'positive' ability. By averaging all six scores a total score was calculated for 22 CI users.

Nijmegen Cochlear Implant Questionnaire

The Nijmegen Cochlear Implant Questionnaire (NCIQ) is a validated, CI specific health-related QoL (HRQoL) instrument (Hinderink, Krabbe, and Van Den Broek 2000). The questionnaire is composed of three categories with six domains: Physical functioning: *sound perception-basic, sound perception-advanced, speech production;* Social functioning: *activity, social functioning;* Psychological functioning: *self-esteem.* The six domains of the NCIQ include ten statements with a five-point response scale. Scores per domain could range from 0 (worst) to 100 (best). A total score was calculated by averaging the scores from all six domains in 22 CI users.

Cochlear Implant Functioning Index

The third questionnaire was the Cochlear Implant Functioning Index (CIFI), a tool to assess the auditory-related functioning of CI users (Coelho et al. 2009). The CIFI was scientifically translated to Dutch by the University of Groningen Language Center, and was further revised by an audiologist, Ear-, Nose- and Throat surgeon and audiology scientists. This questionnaire scores five fields of auditory functioning: 1) *reliance on visual assistance*, 2) *telephone use*, 3) *communication at work*, 4) *'hearing' in noise*, 5) *hearing in groups*, and 6) *hearing in large room settings*. The third field *communication at work* was excluded, because eight out of 22 (36%) respondents were unemployed, making this item not informative for this specific study population. We used total scores ranging from 0 (worst) to 19 (best functioning) in 22 CI users.

Speech, Spatial and Qualities Questionnaire

The Speech, Spatial and Qualities of hearing scale (SSQ) is a validated environmental and spatial hearing questionnaire (Gatehouse and Noble 2004). The Dutch translated version 3.1.2 (2007) was used in this study. The SSQ was developed to quantify the abilities, in particular for speech perception and spatial hearing, in hearing-impaired people and CI users. The questionnaire is composed of three domains: *Speech, Spatial* and *Qualities*. The self-perceived everyday hearing ability is rated with a score between 0 (least) to 10 (maximum ability). A total score was calculated by averaging the scores of all domains in 16 CI users who filled this questionnaire entirely.

Word recognition

Word recognition scores were gathered by trained audiologists during the regular post-implantation outpatient visits as a measure of speech perception (Bosman and Smoorenburg 1995). In the test, meaningful consonant-vowel-consonant words were presented in quiet at 65 and 75 dB SPL (free field) in an audiometry booth. In a list of twelve words, the ratio of correctly repeated phonemes to the total number of phonemes presented was used to calculate a percent correct score. These scores were available for 19 participants at 65 dB SPL and for 20 participants at 75 dB SPL.

Statistical analysis

Spearman's correlation coefficient was used to evaluate the relationships between the scores from DMBQ, NCIQ, SSQ and the word-recognition test. Statistical analyses were processed in Predictive Analytic Software (PASW) software package version 18.0. A level of $p<0.05$ (two tailed) was considered significant.

RESULTS

Enjoyment of music

Figure I shows the satisfaction with listening to music through a CI. A majority of EDLI CI users who answered this section of the DMBQ (12 out of 20 CI users; 60%) rated the sound of music as pleasant.

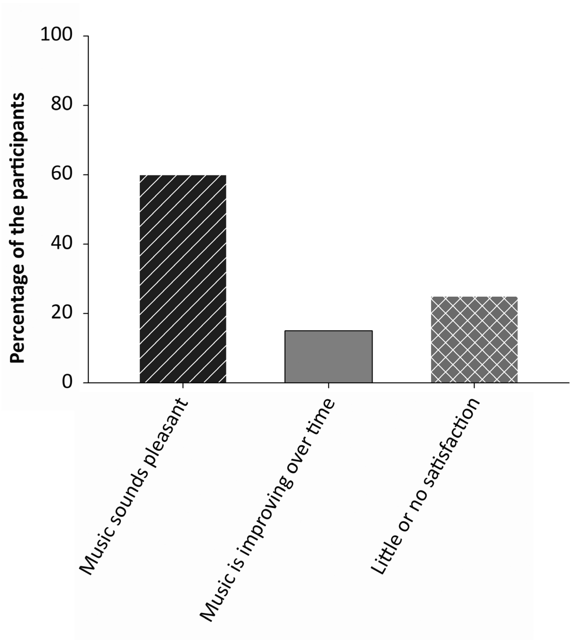


FIGURE 1. The satisfaction with listening to music in 3 categories of the DMBQ in 20 CI users.

Self-perceived quality of music

Figure II shows the scores of the self-perceived quality of music. The mean scores of 22 CI users ranged from 42 to 68 (within the range of 0 to 100), with standard deviations ranging from 23 to 30. The total score was on the positive side of the scale (i.e., larger than 50) with a mean of 56 and a standard deviation of 19.

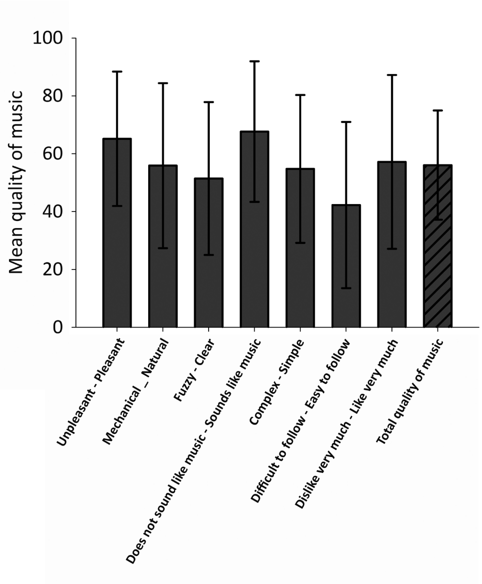


FIGURE 2. The self-perceived quality of music of DMBQ averaged across all 22 participants. The error bars denote 1 standard deviation.

Self-reported perception of the elements of music

Figure III shows the scores of the self-reported perception of the elements of music. A majority indicated to be able to follow the lyrics (18 out of 22 CI users; 82%), recognize the instruments (15 out of 22 CI users; 68%) and follow the melody (13 out of 22 CI users; 59%). The ability to differentiate between singing and speaking was scored negatively in general (16 out of 22 CI users; 73 %).

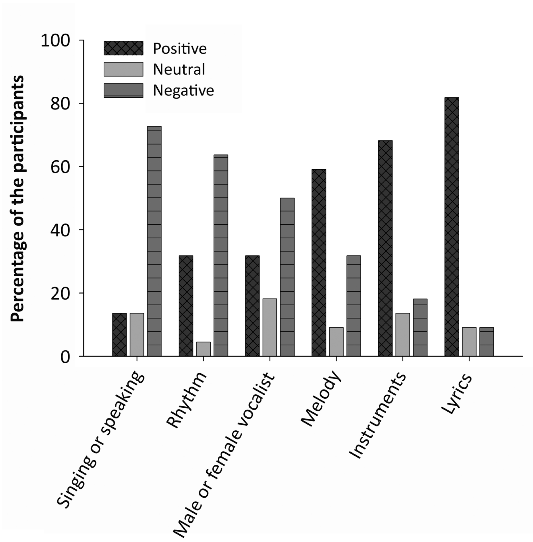


FIGURE 3. The self-reported perception of the elements of music of DMBQ in 22 CI users.

Correlations between DMBQ measures and NCIQ, CIFI and SSQ

Table II shows the correlations between the scores of the DMBQ measures and the NCIQ, CIFI and SSQ. The total NCIQ scores ranged from 44 to 92 (within the range of 0 to 100, best HRQoL) with a mean of 72 in 22 CI users. The total CIFI scores ranged from 4 to 19 (within the range of 4, worst, to 19, best auditory related functioning) with a mean of 11 in 22 CI users. The total SSQ scores ranged from 0.6 to 7.6 (within the range of 0 to 10, best hearing related functioning) with a mean of 4.4 in 16 CI users. No significant correlations were shown between the DMBQ measures and the NCIQ, CIFI, and SSQ scores.

Correlations between DMBQ measures and word recognition scores

Table III shows the correlation analysis between the DMBQ measures and the word recognition scores, ranging from 0 to 95 and a mean of 59%. No significant correlations were observed.

TABLE II: Correlations between all DMBQ measures and the total scores of NCIQ (left column), CIFI (middle column) and SSQ (right column).

	NCIQ	**CIFI**	**SSQ**
Satisfaction with listening to music	r = -0.174	r = -0.311	r = 0.470
	p = 0.462	p = 0.182	p = 0.090
	N = 20	N = 20	N = 14
Self-perceived quality of music	r = -0.007	r = -0.237	r = 0.377
	p = 0.974	p = 0.289	p = 0.150
	N = 22	N = 22	N = 16
Perception of the elements of music	r = 0.179	r = 0.079	r = 0.371
	p = 0.427	p = 0.727	p = 0.157
	N = 22	N = 22	N = 16

TABLE III: Correlations between all DMBQ measures and the word recognition in quiet scores measured at 65 and 75 dB SPL.

	Word recognition 65 dB (%)	**Word recognition 75 dB (%)**
Self-perceived quality of music	r = - 0.194	r = - 0.050
	p = 0.425	p = 0.843
	N = 19	N = 20
Satisfaction with listening to music	r = 0.107	r = - 0.010
	p = 0.672	p = 0.968
	N = 18	N = 18
Perception of the elements of music	r = 0.242	r = 0.301
	p = 0.319	p = 0.197
	N = 19	N = 20

DISCUSSION

The present study evaluated the self-perceived enjoyment and perception of music in the EDLI adult CI users, whose onset of severe hearing loss was at six years of age or younger, and who were implanted at 16 years of age or older. Due to the potentially negative factors, such as less-than complete language development due to early onset of hearing loss and a delay between the onset of severe hearing loss and implantation, this population has historically been not strong candidates for CI implantation. As a result, while implantation in this group has now become more common, very limited knowledge on their music perception is available. The motivation for this study was, therefore, to comprehensively and systematically investigate music-related outcomes of implantation in this understudied group of CI users. We had hypothesized that, unlike the post-lingually deafened CI users (Boucher and Bryden 1997; Drennan and Rubinstein 2008; Gfeller et al. 2000b; Looi and She 2010; Migirov, Kronenberg, and Henkin 2009; Fuller et al. 2012), this group may enjoy music perception, based on previous studies with EDEI CI users. We had further hypothesized, based on post-lingual CI studies, that the enjoyment and perception of music could be correlated with other outcome factors such as the self-reported quality of life, the self-perceived hearing performance and the behaviorally measured word recognition scores (Fuller et al. under revision; Lassaletta et al. 2007; Lassaletta et al. 2008).

Self-perceived enjoyment and perception of music

The results from the music questionnaire on self-perceived enjoyment and perception of music showed that the majority of the EDLI CI users found music to sound pleasant. Additionally, the quality of music was also rated on the positive side of the scale. These observations reconfirm the findings of former studies that showed both the EDEI and EDLI CI users report high satisfaction with listening to music (Migirov, Kronenberg, and Henkin 2009; Eisenberg 1982). However, both satisfaction and quality ratings within these populations are in contrast to the reports of post-lingually deafened CI users, who showed dissatisfaction and lack of enjoyment with music (Boucher and Bryden 1997; Drennan and Rubinstein 2008; Gfeller et al. 2000b; Fuller et al. 2012; Fuller et al. under revision; Looi and She 2010; Migirov, Kronenberg, and Henkin 2009). Several interpretations are possible for the differences observed in music appreciation by EDLI and post-lingually deafened CI users. Firstly, the EDLI CI users might have a different reference point to judge the quality of music with respect to definitions such as complex or simple, or mechanical or natural, due to an underdeveloped acoustical music memory. This situation could be further intensified by years of music listening without proper feedback, and/or with a different modality of listening, such as the tactile representation of music. For example, a song could sound natural to an EDLI but mechanical to a NH person listening to CI-simulations, due to the different states of the auditory exposure and memories of individuals. As a result, different listener groups may be making their music judgment using different standards and reference points (Mitani et al. 2007; Eisenberg 1982; Trehub, Vongpaisal, and Nakata 2009). Moreover, the additional benefit of the implant for music perception compared to the music perception during the period of deafness using a hearing aid may also differ. This would be most likely in the form of more temporal cues and vibrations (Eisenberg 1982). Concluding, the positive self-perceived enjoyment and perception of music in the EDLI group indicates that music could be addressed as an extra factor for implant candidacy of early-deafened clinical populations.

Self-reported perception of the elements of music

The results from the music questionnaire on self-reported perception of the elements of music showed that EDLI CI users indicated to be best able to follow the lyrics and the melody of songs, and to differentiate between musical instruments. They reported that following the rhythm and differentiating between singing and speaking and a male or female vocalist was most difficult. These findings are surprising as they are in contrast to the self-reported perception of these elements of music in post-lingually deafened CI users (Fuller et al. under revision). Post-lingually deafened CI users report to perceive rhythm best, followed by melody and instrument recognition. These self-reports are consistent with the behaviorally tested perception of the elements of music by post-lingually deafened CI users, which show rhythm to be perceived best and melodies worst (Gfeller et al. 2010; Gfeller et al. 2008; Kong et al. 2009; Galvin, Fu, and Nogaki 2007; Galvin, Fu, and Shannon 2009; Gfeller et al. 2005; Nimmons et al. 2008). Based on these findings in post-lingually deafened CI users and keeping in mind the techniques that the CI uses to process sounds leading to loss of fine temporal information, one would expect the EDLI group to be able to follow the rhythm best and not to be able to differentiate the instruments. Although the behavioural perception of music of EDLI CI users has not been studied yet, we might, with some caution, conclude on the basis of the comparison with post-lingually deafened CI users and the findings of our study, that the self-reported perception of the elements of music of EDLI CI user may not be in accordance with the expected behavioural scores. Again this may be explained by a possibly different interpretation of rhythm or melody in EDLI CI users compared to post-lingually deafened based on different reference points. To gain more insight in the differences between early and post-lingually deafened groups, and between self-reported and behaviorally measured music perception, behavioral or objective tests need to be conducted in EDLI CI users to validate this hypothesis.

Correlations between DMBQ and NCIQ, CIFI, SSQ and word recognition

Based on the findings in post-lingually deafened CI users (Fuller et al. under revision; Lassaletta et al. 2007; Lassaletta et al. 2008), we had hypothesized that higher enjoyment and better perception of music would be correlated with higher QoL, better everyday hearing ability and better word recognition. As no such correlations were shown, the results did not support this hypothesis. The different findings between these groups may imply that the self-perceived enjoyment and perception of music is not a significant contributing factor to the QoL and hearing-related functioning for the EDLI users, unlike for the post-lingually deafened CI users. QoL is a complex entity that depends on many factors in life, factors that are probably not all taken into account in this study and that might differ between different CI populations. For example, the gain in speech perception, which can be substantial, might have a larger contribution to the quality of life in EDLI than in postlingually deafened, reducing the potential effects of music-related factors. Also, the absence of correlations between the perception of music and the other outcomes could be caused by the different interpretation of music by EDLI CI users, as mentioned above. A last factor that should be discussed for better interpretation of our data is the number of participants of the current study (N=22). Although the group of EDLI CI users is a slowly expanding group, currently, this clinical population is still small worldwide. Reflecting this general limitation, the number of participants in this study might have been insufficient to find significant correlations between the perception and enjoyment of music and the quality of life. In comparison, with the larger groups of postlingually deafened CI users, Lasaletta et al. (2007) and Fuller et al. (in revision) did show such correlations in 52 and 98 post-lingually deafened CI users, respectively. Therefore, further research needs to be conducted in the growing group of EDLI CI users to gain more insight in the outcome measures including music perception and enjoyment, both subjectively and behaviorally tested.

CONCLUSION

Concluding, overall results of the study showed that EDLI CI users enjoy the perception of music, rate the quality of music high and are satisfied with listening to music using their CIs. Traditionally the criteria for implantation have excluded early-deafened adults and adolescents, because the long duration of auditory deprivation, the minimal exposure to important sounds, such as speech and music, and the underdeveloped auditory memory may make the brain unable to adapt to the implant, preventing effective use of it (Luxford 1989). In the last decade, however, while the outcomes in EDLI CI users tend to be poorer for speech perception compared to EDEI, an improvement in speech perception due to CIs has consistently been shown with this population (Mallinckrodt et al. ; Klop et al. 2007; Schramm, Fitzpatrick, and Séguin 2002; Dowell et al. 2002; Waltzman, Roland, and Cohen 2002; Waltzman and Cohen 1999). Complementing these earlier studies that showed a speech perception benefit, the present study showed high enjoyment and satisfaction with listening to music post-implantation. These new findings of the present study may give additional support for cochlear implant candidacy of (well-selected) early-deafened individuals.

Acknowledgment

The authors thank Prof. Dr. Sieneke Goorhuis-Brouwer for her help with defining the speech language acquisition periods in Dutch. The authors also acknowledge those who worked on The Speech, Spatial and Qualities of Hearing Scale (developed by William Noble (University of New England, Australia) and Stuart Gatehouse (MRC Institute of Hearing Research, Scotland), translated by Liesbeth Royackers (ExpORL, K.U. Leuven, Belgium), and this translation was evaluated by Sophia Kramer (VU MC, Amsterdam, The Netherlands), Wouter Dreschler (AMC, Amsterdam, The Netherlands), Hans Verschuure (ErasmusMC, Rotterdam, The Netherlands),William Damman (AZ St. Jan, Brugge, Belgium), Astrid van Wieringen (ExpORL, K.U.Leuven, Belgium), and Heleen Luts (ExpORL, K.U.Leuven, Belgium).

REFERENCES

- Blamey, P., F. Artieres, D. Başkent, F. Bergeron, A. Beynon, E. Burke, N. Dillier, et al. 2013. Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiology & Neuro-Otology* 18 (1): 36-47.
- Bosman, A. J., and G. F. Smoorenburg. 1995. Intelligibility of dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology* 34 (5): 260-84.
- Boucher, R., and M. P. Bryden. 1997. Laterality effects in the processing of melody and timbre. *Neuropsychologia* 35 (11) (Nov): 1467-73.
- Coelho, D. H., P. E. Hammerschlag, Y. Bat-Chava, and D. Kohan. 2009. Psychometric validity of the cochlear implant function index (CIFI): A quality of life assessment tool for adult cochlear implant users. *Cochlear Implants International* 10 (2): 70-83.
- De Raeve, L. 2010. A longitudinal study on auditory perception and speech intelligibility in deaf children implanted younger than 18 months in comparison to those implanted at later ages. *Otology & Neurotology* 31 (8) (Oct): 1261-7.
- Dowell, R. C., S. J. Dettman, K. Hill, E. Winton, E. J. Barker, and G. M. Clark. 2002. Speech perception outcomes in older children who use multichannel cochlear implants: Older is not always poorer. *The Annals of Otology, Rhinology & Laryngology.Supplement* 189 (May): 97-101.
- Drennan, W. R., and J. T. Rubinstein. 2008. Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of Rehabilitation Research and Development* 45 (5): 779-89.
- Eisenberg, L. S. 1982. Use of the cochlear implant by the prelingually deaf. *The Annals of Otology, Rhinology & Laryngology.Supplement* 91 (2 Pt 3) (Mar-Apr): 62-6.
- Fuller, C., R. Free, B. Maat, and D. Başkent. 2012. Musical background not associated with self-perceived hearing performance or speech perception in postlingual cochlear-implant users. *The Journal of the Acoustical Society of America* 132 (2) (Aug): 1009-16.
- Fuller, C. D., R. H. Free, A. Maat, and D. Başkent. under revision. Correlations of self-reported music perception with quality of life and speech perception in cochlear-implant users.
- Galvin, J. J.,3rd, Q. J. Fu, and G. Nogaki. 2007. Melodic contour identification by cochlear implant listeners. *Ear and Hearing* 28 (3) (Jun): 302-19.
- Galvin, J. J.,3rd, Q. J. Fu, and R. V. Shannon. 2009. Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences* 1169 (Jul): 518-33.
- Gatehouse, S., and W. Noble. 2004. The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology* 43 (2): 85-99.
- Gfeller, K., A. Christ, J. F. Knutson, S. Witt, K. T. Murray, and R. S. Tyler. 2000. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *Journal of the American Academy of Audiology* 11 (7) (Jul-Aug): 390-406.
- Gfeller, K., D. Jiang, J. J. Oleson, V. Driscoll, and J. F. Knutson. 2010. Temporal stability of music perception and appraisal scores of adult cochlear implant recipients. *Journal of the American Academy of Audiology* 21 (1) (Jan): 28-34.
- Gfeller, K., J. Oleson, J. F. Knutson, P. Breheny, V. Driscoll, and C. Olszewski. 2008. Multivariate predictors of music perception and appraisal by adult cochlear implant users. *Journal of the American Academy of Audiology* 19 (2) (Feb): 120-34.
- Gfeller, K., C. Olszewski, M. Rychener, K. Sena, J. F. Knutson, S. Witt, and B. Macpherson. 2005. Recognition of “real-world” musical excerpts by cochlear implant recipients and normal-hearing adults. *Ear and Hearing* 26 (3) (Jun): 237-50.
- Goorhuis-Brouwer, SM, AM Schaerlaekens, and Handboek Taalontwikkeling. 2000. Taalpathologie en taaltherapie bij nederlandssprekende kinderen (handbook language development, language pathology and language therapy in dutch speaking children). *Utrecht, De Tijdstroom*.

- Hilliard, R. E. 2003. The effects of music therapy on the quality and length of life of people diagnosed with terminal cancer. *Journal of Music Therapy* 40 (2) (Summer): 113-37.
- Hinderink, J. B., P. F. Krabbe, and P. Van Den Broek. 2000. Development and application of a health-related quality-of-life instrument for adults with cochlear implants: The nijmegen cochlear implant questionnaire. *Otolaryngology--Head and Neck Surgery : Official Journal of American Academy of Otolaryngology-Head and Neck Surgery* 123 (6) (Dec): 756-65.
- Houston, D. M., and R. T. Miyamoto. 2010. Effects of early auditory experience on word learning and speech perception in deaf children with cochlear implants: Implications for sensitive periods of language development. *Otology & Neurotology : Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology* 31 (8) (Oct): 1248-53.
- Klop, W. M. C., J. J. Briaire, A. M. Stiggelbout, and J. H. M. Frijns. 2007. Cochlear implant outcomes and quality of life in adults with prelingual deafness. *The Laryngoscope* 117 (11): 1982-7.
- Kong, Y. Y., J. M. Deeks, P. R. Axon, and R. P. Carlyon. 2009. Limits of temporal pitch in cochlear implants. *The Journal of the Acoustical Society of America* 125 (3) (Mar): 1649-57.
- Lassaletta, L., A. Castro, M. Bastarrica, R. Perez-Mora, B. Herran, L. Sanz, M. J. de Sarria, and J. Gavilan. 2008. Musical perception and enjoyment in post-lingual patients with cochlear implants. *Acta Otorrinolaringologica Espanola* 59 (5) (May): 228-34.
- Lassaletta, L., A. Castro, M. Bastarrica, R. Perez-Mora, R. Madero, J. De Sarria, and J. Gavilan. 2007. Does music perception have an impact on quality of life following cochlear implantation? *Acta Oto-Laryngologica* 127 (7) (Jul): 682-6.
- Lazard, D. S., C. Vincent, F. Venail, P. Van de Heyning, E. Truy, O. Sterkers, P. H. Skarzynski, et al. 2012. Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: A new conceptual model over time. *PloS One* 7 (11): e48739.
- Looi, V., and J. She. 2010. Music perception of cochlear implant users: A questionnaire, and its implications for a music training program. *International Journal of Audiology* 49 (2): 116-28.
- Luxford, WM. 1989. Cochlear implant indications. *American Journal of Otolaryngology* 10 (2): 95-98.
- Mallinckrodt, EM, R. Hofman, CD Fuller, H. Burger, B. Maat, and RH Free. Evaluation of speech perception, quality of life and daily usage in pre- and perlingually deafened patients; implanted during the postlingual time frame *(in Preparation)*.
- McDermott, H. J. 2004. Music perception with cochlear implants: A review. *Trends in Amplification* 8 (2): 49-82.
- Migirov, L., J. Kronenberg, and Y. Henkin. 2009. Self-reported listening habits and enjoyment of music among adult cochlear implant recipients. *Annals of Otology, Rhinology & Laryngology* 118 (5): 350-5.
- Mitani, C., T. Nakata, S. E. Trehub, Y. Kanda, H. Kumagami, K. Takasaki, I. Miyamoto, and H. Takahashi. 2007. Music recognition, music listening, and word recognition by deaf children with cochlear implants. *Ear and Hearing* 28 (2 Suppl) (Apr): 29S-33S.
- Most, T., H. Shrem, and I. Duvdevani. 2010. Cochlear implantation in late-implanted adults with prelingual deafness. *American Journal of Otolaryngology* 31 (6): 418-23.
- Nakata, T., S. E. Trehub, C. Mitani, Y. Kanda, A. Shibasaki, and E. G. Schellenberg. 2005. Music recognition by japanese children with cochlear implants. *Journal of Physiological Anthropology and Applied Human Science* 24 (1): 29-32.
- Nimmons, G. L., R. S. Kang, W. R. Drennan, J. Longnion, C. Ruffin, T. Worman, B. Yueh, and J. T. Rubenstien. 2008. Clinical assessment of music perception in cochlear implant listeners. *Otology & Neurotology* 29 (2) (Feb): 149-55.
- Niparko, J. K., E. A. Tobey, D. J. Thal, L. S. Eisenberg, N. Y. Wang, A. L. Quittner, N. E. Fink. 2010. Spoken language development in children following cochlear implantation. *Jama* 303 (15): 1498-506.
- Salimpoor, V. N., M. Benovoy, K. Larcher, A. Dagher, and R. J. Zatorre. 2011. Anatomically distinct dopamine

release during anticipation and experience of peak emotion to music. *Nature Neuroscience* 14 (2): 257-62.

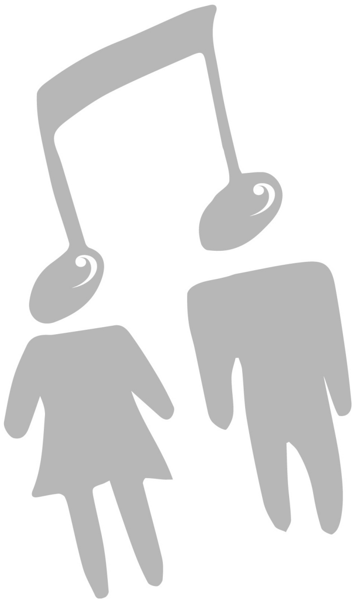
- Santarelli, R., R. De Filippi, E. Genovese, and E. Arslan. 2008. Cochlear implantation outcome in prelingually deafened young adults. A speech perception study. *Audiology & Neuro-Otology* 13 (4): 257-65.
- Schramm, D., E. Fitzpatrick, and C. Séguin. 2002. Cochlear implantation for adolescents and adults with prelinguistic deafness. *Otology & Neurotology* 23 (5): 698-703.
- Trehub, S. E., T. Vongpaisal, and T. Nakata. 2009. Music in the lives of deaf children with cochlear implants. *Annals of the New York Academy of Sciences* 1169 (Jul): 534-42.
- Waltzman, S. B., and N. L. Cohen. 1999. Implantation of patients with prelingual long-term deafness. *The Annals of Otology, Rhinology & Laryngology.Supplement* 177 (Apr): 84-7.
- Waltzman, S. B., J. T. Roland Jr, and N. L. Cohen. 2002. Delayed implantation in congenitally deaf children and adults. *Otology & Neurotology* 23 (3) (May): 333-40.
- Wang, S, B. Liu, R. Dong, Y. Zhou, J. Li, B. Qi, X. Chen, D. Han, and L. Zhang. 2012. Music and lexical tone perception in chinese adult cochlear implant users. *The Laryngoscope* 122 (6): 1353-60.
- Yoshida, H., Y. Kanda, I. Miyamoto, T. Fukuda, and H. Takahashi. 2008. Cochlear implantation on prelingually deafened adults. *Auris Nasus Larynx* 35 (3): 349-52.

Chapter 3

Self-reported music perception is related to quality of life and self-reported hearing abilities in cochlear-implant users

Christina Fuller [1,2] Rolien Free [1,2] Bert Maat [1,2] and Deniz Başkent [1,2]

Under revision



1 Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

2 Graduate School of Medical Sciences (Research School of Behavioural and Cognitive Neurosciences), University of Groningen, The Netherlands.

ABSTRACT

Objective: We hypothesized that cochlear implant (CI) users' music listening habits, music quality ratings and music perception would be related with: 1) quality of life (QoL) and 2) speech perception and hearing ability.

Design: Post-lingually deafened CI participants evaluated themselves in terms of music perception, QoL, and hearing abilities using questionnaires. Additionally, speech perception was behaviorally measured.

Study Sample: Ninety-eight post-lingually deafened CI users.

Results: Music perception after cochlear implantation was significantly related with QoL and self-reported hearing ability.

Conclusions: The findings suggest some relationship between CI user's music perception and self-reported QoL and hearing ability. Music training programs and/or device improvements that improve music perception may also improve QoL and hearing ability in CI users.

INTRODUCTION

Cochlear implants (CIs) are auditory prosthetic devices that restore hearing to individuals with profound to severe sensorineural hearing impairment. CIs are able to provide good levels of speech perception in quiet and a general increase in quality of life (QoL) post-implantation (Faber, Aksel, and Grøntved 2000; Krabbe, Hinderink, and van den Broek 2000; Zhao, Bai, and Stephens 2008). However, music perception and enjoyment are still not satisfactory (Drennan and Rubinstein 2008; Gfeller et al. 2000; Philips et al. 2012).

Perception of the four basic elements in music -pitch, rhythm, melody and timbre- is less accurate and more variable in CI users compared to normal hearing (NH) listeners (Drennan and Rubinstein 2008; McDermott 2004). This discrepancy is partially due to differences between acoustic and electric hearing. CI users' music perception is limited by the coarse spectral resolution (due to the limited number of stimulation sites in the cochlea) and speech processing strategies that retain slowly varying spectro-temporal information but not the spectro-temporal fine structure information (for review, see McDermott 2004). The coarse spectral resolution limits CI users' pitch, melody and timbre perception, where fine structure cues are important (Shannon et al., 2004; Gfeller et al. 2002; Kong et al. 2009; Looi et al. 2008). Only rhythm perception appears to be similar between NH and CI listeners (Gfeller et al. 2007; Kong et al. 2004).

However, the limited *music perception* does not necessarily limit CI users' *music appreciation,* as factors that contribute to music perception and appreciation may be different (Fuller et al. 2013; Gfeller et al. 2000; Gfeller et al. 2008; Gfeller et al. 2010; Lassaletta et al. 2008; Looi et al. 2008; Looi and She 2010; Looi, Gfeller, and Driscoll 2012; Mirza et al. 2003; Wright and Uchanski 2012). Therefore, evaluation of CI outcomes in terms of music should be more comprehensively investigated by evaluating not only behaviorally measured music perception, but also self-reported perception and enjoyment of music. Music is a pervasive art form, an environmental sound and a potent pleasurable stimulus that can positively affect emotional state (Gfeller et al. 2000; Looi, Gfeller, and Driscoll 2012; Salimpoor et al. 2011). Music therapy has been shown to improve QoL in some patient groups (Hilliard 2003; Walworth et al. 2008); such therapy might also have a positive effect for CI users. Therefore, it is important to understand factors that make some CI users appreciate music and others not. Such knowledge would be useful in designing rehabilitation protocols that include music perception and appreciation for CI users.

In addition to having a positive effect on emotional state and QoL, music experience has also been shown to have a positive effect on hearing and speech perception abilities in NH listeners (Parbery-Clark, Skoe, and Kraus 2009; Parbery-Clark et al. 2009). However, musical training and involvement before cochlear implantation did not affect CI users' post-implantation speech perception performance (Fuller et al. 2012). It is possible that other factors related to the CI (e.g., functional spectral resolution) may contribute more strongly to CI outcomes and may have obscured potential music training benefits. As such, CI listeners should not be discouraged to improve their music perception and appreciation, as this may lead to greater CI use, which may lead to better overall performance.

To gain more insight, Lassaletta et al. (2007) and Philips et al. (2012) studied CI users' self-reported perception and enjoyment of music and their association with QoL and speech perception, respectively. Lassaletta et al. used a music questionnaire and a generic QoL questionnaire [Glasgow Benefit Inventory (GBI), which assesses patient benefit after otolaryngological procedures; Robinson, Gatehouse, and Browning 1996] in 52 CI recipients. They found that the self-reported quality of music was correlated with the time spent listening to music with the CI, and with QoL. However, no data on speech perception was collected, and therefore it was unclear how music enjoyment related to speech perception performance. Philips et al. (2012) investigated the relationship between self-reported quality/enjoyment of music and speech perception. Forty CI users answered a newly developed questionnaire on music appreciation and 15 of these participants were subsequently tested for speech perception in quiet and in noise. Music quality and enjoyments were significantly correlated with speech reception thresholds (SRTs) in quiet and in noise. However, as speech perception scores were available only from 15 out of 40 participants (38%), the generalizability of the findings was limited.

The present study investigated potential relationships among music listening habits, self-reported perception of music, QoL, self-reported hearing ability and behaviorally measured speech perception in a large sample (n=98) of post-lingually deafened CI users. We hypothesized that music listening habits, music quality, and music perception would be significantly related with QoL, self-reported hearing ability, and speech perception scores. Questionnaires were used to investigate music listening habits, quality and perception, as well as health-related QoL and hearing ability; behaviorally measured phoneme-in-word recognitions scores were used to quantify speech perception.

MATERIALS & METHODS

Study population

The study population of this study was the same as in Fuller et al. (2012). Two hundred fourteen CI users, selected from patients implanted and/or monitored at the University Medical Center Groningen, were sent three questionnaires. The inclusion criteria were based on: current age (older than 18 years), age at the onset of profound hearing loss (6 years or older to ensure post-lingual deafness; Goorhuis-Brouwer and Schaerlaekens 2000) and more than one year of CI experience. To include as many patients as possible and thus to study a general and representative CI population, etiology and speech perception performance were not used as inclusion criteria. Ninety-eight (46%) replies were received. The demographics of the participants are shown in Table 1. The levels of education refer to the highest completed educational level: low refers to elementary school only, middle refers to middle school or higher, high refers to at least a bachelor's degree. Except for one CI user, all were unilaterally implanted. A comparison was made between the demographics of respondents and non-respondents to ensure that the respondents were indeed a good representation of the larger CI population who were originally sent the questionnaires. Confirming this, no significant differences were observed for age, CI experience, and gender (T-test: t =-1.038, p = 0.301, t = -1,314 p= 0.191, Chi-square-test: χ^2 0.041, p =0.840, respectively).

3

TABLE I: Demographics of the study participants. N refers to number of participants in each table and figure.

Total participants (n)	98
Gender (n)	
Male	39
Female	59
Mean age (y)	65.6 ± 11.9
Level of education	
Lower	12
Middle	67
Higher	14
Mean duration of impaired hearing (y)	37.9 ± 18.6
Mean CI use since implantation (m)	65.7 ± 33.0
Mean CI use per day (h)	15.0 ± 2.6
Hearing aid on the contra-lateral ear	36 (35%)
Implant type (n)	
CI22M[a]	1
CI24R CA[a]	24
CI24R k[a]	5
CI24RE CA[a]	27
CI24R CS[a]	16
HiRes90K Helix[b]	26
Speech processor type (n)	
Esprit3G[a]	31
Freedom[a]	42
Harmony[b]	26
Phoneme recognition in quiet (presented at 65 dB SPL)	65% (std=24%)
Phoneme recognition in quiet (presented at 75 dB SPL)	70% (std=21%)

[a] Cochlear Ltd, Macquarie University, Australia. ACE speech strategy.
[b] Advanced Bionics Corp., California, USA device. HiRes speech strategy.

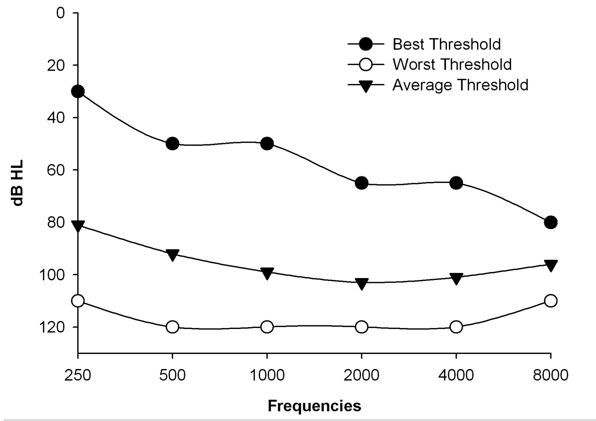


FIGURE 1. Best, worst and average hearing thresholds (across 92 CI participants) measured in the contra-lateral ear before cochlear implantation.

Figure 1 shows the best, average and worst residual acoustic hearing thresholds measured for the contra-lateral ear before implantation. Even though some CI users show useful acoustic hearing at some frequencies, the average thresholds indicate severe hearing loss. To not complicate an already large comprehensive study further, and because the participants were a good representation of typical CI users, it was decided not to additionally analyze the potential effects of residual hearing.

The study was approved by the Medical Ethical Committee of the University Medical Center Groningen. The study was conducting in accordance to the principles expressed in the Declaration of Helsinki. Detailed information about the study was provided to the participants and written informed consent was obtained. Participation was purely voluntary and no financial reimbursement was provided.

Dutch Musical Background Questionnaire

The first questionnaire, the Dutch Musical Background Questionnaire (DMBQ), is a translated and edited version of the Iowa Musical Background Questionnaire (IMBQ) developed by Gfeller et al. (2000)[1]. The questionnaire was translated by a professional translator with assistance from the first author, and was further revised by audiologists, an Ear-, Nose- and Throat surgeon and a psychologist. For the present study only the sections regarding music listening habits, music quality, and perception of basic elements of music were used.

1. Music listening habits

The first part of DMBQ assessed music listening habits. Music listening habits before and after implantation were scored in two items. The first item evaluated the interest in listening

1 Translated by M. Trommelen and C. Fuller.

to music via the statement: *I would describe myself as a person who often chooses to listen to music.* Respondents indicated their agreement with the statement on a one ('strongly disagree') to four ('strongly agree') rating scale. The second item scored the hours spent listening to music per week and was scored on a one to four rating scale: one = 0 to 2 hours, two = 3 to 5, three = 6 to 8 hours, and four = more than 9 hours. Adding the scores from the two items, two cumulative scores were calculated for music listening habits: one pre-implantation and one post-implantation. The total score thus ranged from 2 to 8. Note that not all 98 participants filled all sections of all questionnaires; therefore the numbers of participants (N) for specific sections will be indicated explicitly in text, figures, and tables. Seventy-four participants completed this part of the DMBQ.

2. Subjective quality of music

The second part of DMBQ assessed music quality with the CI. The recipients were asked to indicate how music sounds under the best conditions with their CI. Seven visual analogue scales (VASs), each ranging from 0 (worst) to 100 (best), were used. The extremes of each VAS were coupled to opposite adjective descriptors (*unpleasant-pleasant*, *mechanical-natural, fuzzy-clear, does not sound like music-sounds like music, complex-simple, difficult to follow-easy to follow, dislike very much-like very much*). An overall mean score between 0 and 100, calculated by averaging across the seven scales, was used to quantify the subjective quality of music. Ninety-seven participants completed this section.

3. Elements of music

The third part of DMBQ investigated the ability to perceive the elements of music (rhythm, melody and timbre), to differentiate vocalists, and to follow the lyrics of a song. The questions were scored on a seven-point scale ranging from 1 (never) to 7 (always). The values 1 to 3 thus indicated a 'negative' ability, 4 a 'neutral' ability and 5 to 7 a 'positive' ability. The specific questions were:

1. *Can you hear the difference between singing and speaking?*
2. *Are you able to differentiate between a male and a female vocalist?*
3. *Are you able to follow the rhythm of a music piece?*
4. *Are you able to recognize the melody of a music piece?*
5. *Are you able to differentiate the instruments in a piece of music?*
6. *Can you follow the lyrics of a song?*

A total mean score between 1 and 7 was calculated by averaging the scores from all six questions used to quantify the ability to perceive music elements. Eighty-seven participants completed this section.

Nijmegen Cochlear Implant Questionnaire

The second questionnaire, the Nijmegen Cochlear Implant Questionnaire (NCIQ), is a validated CI specific, health-related QoL questionnaire (Hinderink, Krabbe, and Van Den Broek 2000). The NCIQ has three categories in which six domains are allocated: *physical functioning (sound perception-basic, sound perception-advanced, and speech production*), *social functioning* (*activity, social functioning*), *and psychological functioning* (*self-esteem*). Scores range from 0 (worst) to 100 (best) per domain. A total mean score between 0 and 100 was calculated by averaging across all six domains. Ninety-two participants completed the NCIB.

Speech, Spatial and Qualities Questionnaire

The third questionnaire, the Speech, Spatial and Qualities of hearing scale (SSQ)[2], is a measure of hearing performance, validated for hearing-impaired listeners and CI users (Gatehouse and Noble 2004). The Dutch translated version 3.1.2 (2007) was used in this study. The SSQ covers three domains of hearing: *speech*, *spatial*, and other *qualities*. Respondents rated themselves with scores varying from 0 (worst) to 10 (best). A total mean score between 0 and 10 was calculated by averaging scores across all three domains. Seventy-three participants completed the SSQ.

Recognition of phonemes in words

Recognition of phonemes in words was measured during the regular outpatient visits by trained clinical audiologists. Meaningful and commonly used consonant-vowel-consonant words were presented at 65 and 75 dB SPL in quiet (Bosman and Smoorenburg 1995). One list of twelve words was played per dB-level in free field. A list was presented using an audiometer (Equinox 2.0 from Interacoustics; Lanarkshire, Scotland) via a power amplifier (AP 12 Ritmton; Samsun, Turkey) with the patient facing the speaker (DALI, Interacoustics; Lanarkshire, Scotland) at 2.5 meter in an audiometry booth. The ratio of correctly repeated phonemes to the total number of phonemes presented was used to calculate the percent correct score. Speech perception scores were available for 71 participants at 65 dB SPL and for 72 participants at 75 dB SPL.

Statistics

Multiple linear regression analyses were used to compare results from NCIQ, SSQ and speech perception to the music measures from the DMBQ. A level of $p < 0.05$ (two tailed) was considered significant. Statistical analyses were run using SPSS 20.

2 Developed by William Noble (University of New England, Australia) and Stuart Gatehouse (MRC Institute of Hearing Research, Scotland), translated by Liesbeth Royackers (ExpORL, K.U.Leuven, Belgium) and this translation was evaluated by Sophia Kramer (VU MC, Amsterdam, The Netherlands), Wouter Dreschler (AMC, Amsterdam, The Netherlands), Hans Verschuure (Erasmus MC, Rotterdam, The Netherlands), William Damman (AZ St. Jan, Brugge, Belgium), Astrid van Wieringen (ExpORL, K.U.Leuven, Belgium) and Heleen Luts (ExpORL, K.U.Leuven, Belgium)

RESULTS

Dutch Musical Background Questionnaire

Music listening habits

Figure 2 shows the results of the music listening habits part of the DMBQ. The upper panel shows the interest in listening to music, the middle panel the time spent listening to music per week, and the bottom panel the total scores of the music listening habits before and after implantation ranging from 2 (worst) to 8 (best). Figure 2 shows a significant decline in music listening habits after implantation, reflected in all three panels (all $p < 0.000$, from top to bottom panels, z -5.008, z -5.738, z -5.673, respectively, by Wilcoxon signed rank test). The interest in listening to music and the hours spent listening to music (top and middle panels, respectively) were significantly correlated before ($r = 0.538$, $p < 0.001$) and after implantation ($r = 0.567$, $p < 0.001$).

3

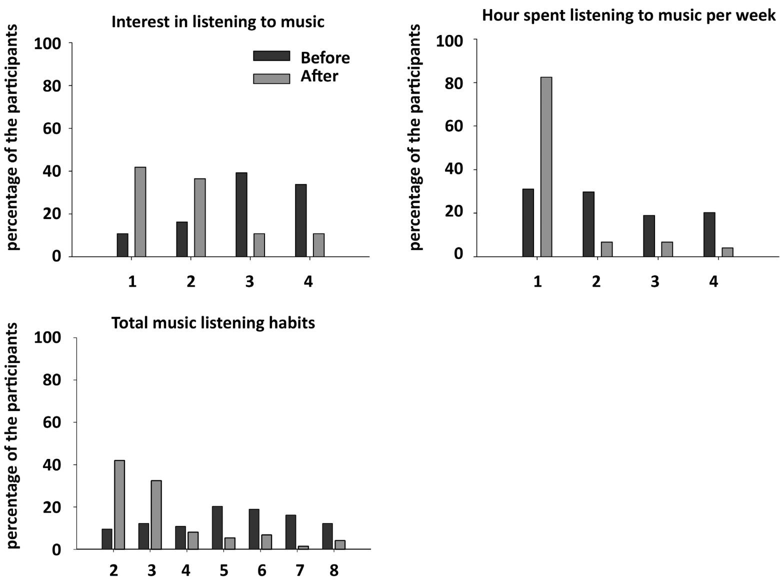


Figure 2: Self-reported music listening habits before and after implantation. The upper right panel shows the results from the first item, the interest in listening to music expressed via agreement with the statement: I would describe myself as a person who often chooses to listen to music. The upper left panel shows the time spent listening to music per week. The bottom right panel shows the total scores for music listening habits, calculated by adding the scores of the two top panels. The total scores thus ranged from 2 (minimum music listening habits) to 8 (maximum music listening habits).

Subjective quality of music

Figure 3 shows the average results (across all participants) for the subjective quality of music with the CI on a 0 (worst) to 100 (best) scale, for the individual adjectives (orange bars), as well as the total quality of music (blue bar). All mean scores were below 50, on the negative half of the scale.

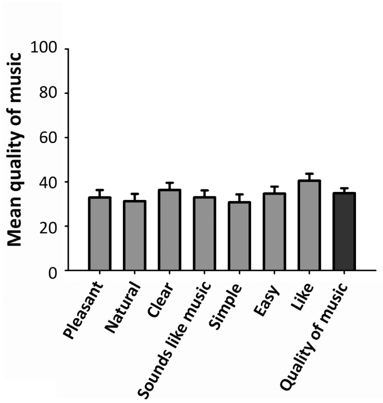


FIGURE 3: The self-reported quality of music, scored between 0 (worst) and 100 (best) and shown separately for the seven descriptor pairs. The combined total score, averaged across the seven scales, is shown by the rightmost column. The error bars denote one standard error.

Elements of music

Figure 4 shows the results of the subjective perception of the elements of music, reported in percentages of the participants. The majority of the respondents reported to be able to differentiate between singing and speaking (58%) and between a female or male vocalist (53%). From the structural elements of music (i.e. rhythm, melody and timbre) the CI recipients reported to be best able to recognize rhythm. Forty-four percent of the recipients were able to follow the rhythm, 23% recognize the melody and 15% identify musical instruments. The recipients reported the lyrics as the most problematic of these elements to follow. None (0%) of the CI users was *always* able to follow the lyrics and 44% were *never* able to follow the lyrics.

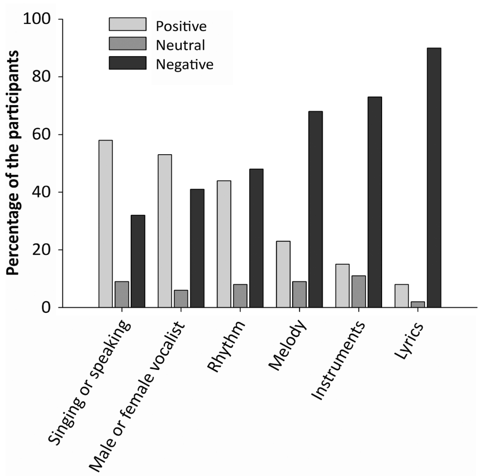


FIGURE 4: The differentiation and recognition of the elements of music, shown in percentages of the respondents who reported a positive, neutral, or negative ability.

NCIQ questionnaire

Table II shows the mean scores per domain and the total score for the NCIQ. There was a wide range in total NCIQ scores, ranging from 20 to 88, with a mean of 62, on a 0 (minimum health-related QoL) to 100 (maximum health-related QoL) scale.

TABLE II: Mean scores and standard deviations of the domains and total scores of the NCIQ (between 0 and 100).

NCIQ	Mean (standard deviation)
Sound perception basic	55 (21)
Sound perception advanced	47 (19)
Speech production	76 (16)
Self esteem	63 (17)
Activity limitations	65 (20)
Social interactions	65 (16)
Total NCIQ	62 (15)

SSQ questionnaire

Table III shows the scores per domain and the total score for the SSQ. The total SSQ scores ranged from 0 to 7.6, with a mean of 3.5, on a 0 (no hearing ability) to 10 (maximum hearing ability) scale.

TABLE III: Mean scores and standard deviations of the domains and total scores of the SSQ (between 0 and 10).

SSQ	Mean (standard deviation)
Speech	3.2 (1.8)
Spatial	3.0 (2.1)
Qualities of hearing	3.9 (1.9)
Total SSQ	3.4 (1.7)

Speech perception scores

Mean recognition of phonemes in words was 54% correct (range: 0-97) at 65 dB SPL and 67% correct (range: 0-97) was at 75 dB SPL.

Regression analyses

Because not all participants completed all questionnaires, separate multiple linear regression analyses were performed between the DMBQ music measures and the NCIQ, SSQ, and speech measures (Table IV). Significant relationships were observed between the DMBQ and the NCIQ and SSQ measures ($p < 0.05$ in both cases). Because the number of subjects differed across measures, it was not possible to strictly correct for family-wise error associated with multiple comparisons. However, using a Bonferroni adjustment to the significance level (0.05/4 = 0.0125), the significant relationships persisted between the DMBQ and the NCIQ and SSQ measures. Only the elements of music was found to contribute significantly to the regression ($p = 0.001$ in both cases). There was no significant relationship between either speech measure and the music measures ($p > 0.05$ in both cases).

TABLE IV: Multiple linear regressions between CI outcome measures and DMBQ measures.

	Regression fit			Pre-CI		Post-CI		Quality		Elements	
	n	r	p	t	p	t	p	t	p	t	p
NCIQ	67	0.50	0.001	-0.59	0.558	-0.23	0.822	0.44	0.663	3.54	0.001
SSQ	55	0.50	0.007	-0.94	0.351	-1.36	0.160	0.22	0.830	3.70	0.001
Speech 65	51	0.31	0.303	-0.51	0.611	0.23	0.820	0.78	0.442	1.11	0.771
Speech 75	52	0.34	0.209	0.37	0.713	-0.10	0.924	0.81	0.425	1.54	0.130

DISCUSSION

In the present study, self-reported music perception (DMBQ) in post-lingually deafened CI users was investigated and compared to outcome measures in terms of self-reported QoL (NCIQ), self-reported hearing ability (SSQ), and behaviorally measured speech perception (phoneme-in-word recognition at 65 and 75 dB SPL). We hypothesized that listening habits, better quality, and perception of music would be associated with the NCIQ, SSQ, and speech perception. While significant relationships were found between the music measures and the NCIQ and SSQ, these were largely driven by perception of elements of music; no significant relationships were observed between the DMBQ and speech perception.

Note that the same study population was used as in Fuller et al. (2012), presented in Chapter 6. The hypotheses of these studies were different. In Fuller et al. (2012), we hypothesized that formal music training before implantation (measured with different questions of the DMBQ) would affect QoL, self-reported hearing ability, and speech perception. In this study, the music measures were not sensitive to formal music training, and represented general music listening experience, quality, and perception. Because the hypotheses were different, and to present the data more clearly, the two studies are presented in different chapters of the thesis and were submitted as different papers.

Music factors

In accordance with literature, a decline in the music listening habits after implantation has been previously reported in post-lingually deafened CI users (Gfeller et al. 2000; Lassaletta et al. 2007; Lassaletta et al. 2008; Looi and She 2010; Migirov, Kronenberg, and Henkin 2009; Mirza et al. 2003; Philips et al. 2012). In this and these previous studies, music quality with the CI was rated negatively in general.

For music perception with the CI, participants reported that they were most able to differentiate between singing and speaking and between a female and a male vocalist. The latter was scored even more positively than the ability to follow the rhythm. This is surprising because the differentiation between a female or a male vocalist depends mostly on voice pitch and to a lesser extent on timbre; CI users' voice gender recognition has been shown to be more difficult than rhythm identification (Fu, Chinchilla, and Galvin 2004; Gfeller et al. 2007). Consistent with our findings, Philips et al. (2012), using questionnaires, reported that 53% of CI subjects indicated they were able to distinguish between male and female voices (compared to 58% in this study), while only 30% were able to follow the rhythm (compared to 44% in this study). Thus, while CI users seem to be able to follow simple rhythms in behavioral studies, they subjectively report they are unable to follow the rhythm in musical pieces. This difference could be due to the 'rhythm-only excerpts' used in behavioral studies compared to the overall perception of rhythm music encountered in daily life (Drennan and Rubinstein 2008; Gfeller et al. 2007; Kong et al. 2004; Won, Drennan, and Rubinstein 2007).

Considering the basic elements of music - rhythm, timbre and melody- the order of rating for the ability to perceive them was as expected. Rhythm was reported to be perceived best, followed by timbre and subsequently by melody. This is consistent with the results of both behavioral studies and subjective questionnaires (Gfeller et al. 2007; Philips et al. 2012). It was somewhat surprising that the present participants rated lyric perception in music to be most problematic, with 44% reporting that they were *never* able to follow the lyrics.

Previous CI studies have reported that lyrics were beneficial for perception and recognition of music (Gfeller et al. 2002a; Leal et al. 2003). Again, being able to follow the lyrics of short musical excerpts used for behavioral testing may be different than a more general perception of lyrics in music encountered in everyday life. In some ways, the ability to follow lyrics is akin to the intelligibility of speech in music. Consistent with our findings, speech intelligibility in background music has been observed to be poorer in CI users than in NH listeners (Eskridge et al. 2012).

Music versus quality of life

The perception of music elements was the only component of the DMBQ that was predictive of QoL, as measured with the NCIQ. Music listening habits before/after implantation and music quality were not predictive of QoL after or the quality of the sound of music was found. Fuller et al. (2012) similarly found no significant relationship between musical background before implantation and health-related QoL in the same groups of subjects.

There is some agreement between the present findings and those from previous studies. Lassaletta et al. (2007) showed a significant positive association between music listening habits, music quality, and QoL in 52 adult CI users, using different questionnaires than in the present study. Zhao et al. (2008) found that improvement in QoL was related to different variables for individual CI subjects. In 38% of CI subjects, speech communication was a key determinant of QoL; in 25% of CI subjects, music perception was and in three out of twelve subjects improved music was a key determinant for QoL. Music perception and Qol may both be influenced by device-related factors (e.g., electrode placement, quality of electrode-nerve signal transmission, etc.) and/or patient-related factors (etiology, health of the spiral ganglia, cognitive elements, etc.).

Music versus hearing abilities and speech perception

Perception of music elements was the only component of the DMBQ that was predictive of hearing ability, as measured with the SSQ. Speech perception (as measured by phoneme recognition in quiet at 65 and 75 dB) was not significantly related to any of components of the DMBQ. The lack of relation between speech and music perception may be due to spectral resolution. While four spectral channels may be adequate for speech recognition in quiet, many more channels are required for music perception (Shannon et al., 2004). Thus, good speech performers may have rated music perception poorly, or that their music listening habits involved less time than speech perception, which is a more constant listening demand. Speech recognition in noise or pitch-based speech perception (e.g., voice gender categorization, vocal emotion recognition, etc.) may have been more strongly related to music perception. Philips et al. (2012) reported that enjoyment of music and quality were correlated with CI users' speech reception thresholds in quiet and in noise. Won and colleagues (2007; 2010) found that word recognition in quiet was related to specific music elements of melody, timbre, and pitch, suggesting that improvements in CI signal processing that improve speech perception might also improve music perception, and vice versa.

Improved music perception via music training may benefit speech perception, as music experience has been shown to relate to NH listeners' speech performance (Parbery-Clark et al. 2009). The results of the present study have important implications, as some aspects of music perception were strongly linked to QoL and self-reported hearing abilities. CIs were originally developed and optimized for speech perception. Developing CI technology to improve music perception may have a strong positive effect on QoL. Other benefits may also come from improved music perception (e.g. better performance in challenging environments, better perception of important pitch cues in speech, etc.). With improved CI technology and/or music training, improvements in music perception may have profound effects on CI outcomes.

Acknowledgements

We would like to show our gratitude to Astrid van Wieringen, Wouter Dreschler, Emmanuel Mylanus for help with the QoL questionnaires, Virginia Discroll and Kate Gfeller for help with IMBQ, and Rosemarie Arnold for her suggestions on the questionnaires in general. We would like to thank all the students from 'the aquarium' for their packing skills. The second author is supported by an otological/neurotological stipendium from the Heinsius-

Houbolt Foundation. The last author is supported by a Rosalind Franklin Fellowship from the University Medical Center Groningen, University of Groningen, and the VIDI grant 016.096.397 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw). The study is part of the research program of our department: Healthy Aging and Communication. There is no conflict of interest regarding this manuscript.

REFERENCES

- Bosman, A. J., and G. F. Smoorenburg. 1995. Intelligibility of dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology* 34 (5): 260-84.
- Drennan, W. R., and J. T. Rubinstein. 2008. Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of Rehabilitation Research and Development* 45 (5): 779-89.
- Eskridge, E. N., J. J. Galvin 3rd, J. M. Aronoff, T. Li, and Q. J. Fu. 2012. Speech perception with music maskers by cochlear implant users and normal-hearing listeners. *Journal of Speech, Language, and Hearing Research : JSLHR* 55 (3) (Jun): 800-10.
- Faber, E., M. Aksel, and C. Grøntved. 2000. Cochlear implantation and change in quality of life. *Acta Oto-Laryngologica* 120 (543): 151-3.
- Fu, Q. J., S. Chinchilla, and J. J. Galvin. 2004. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology* 5 (3): 253-60.
- Fuller, C., R. Free, B. Maat, and D. Başkent. 2012. Musical background not associated with self-perceived hearing performance or speech perception in postlingual cochlear-implant users. *The Journal of the Acoustical Society of America* 132 (2) (Aug): 1009-16.
- Fuller, C., L. Mallinckrodt, B. Maat, D. Başkent, and R. Free. 2013. Music and quality of life in early-deafened late-implanted adult cochlear implant users. *Otology & Neurotology* 34 (6) (Aug): 1041-7.
- Galvin, J. J.,3rd, Q. J. Fu, and G. Nogaki. 2007. Melodic contour identification by cochlear implant listeners. *Ear and Hearing* 28 (3) (Jun): 302-19.
- Gatehouse, Stuart, and William Noble. 2004. The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology* 43 (2): 85-99.
- Gfeller, K., A. Christ, J. F. Knutson, S. Witt, K. T. Murray, and R. S. Tyler. 2000. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *Journal of the American Academy of Audiology* 11 (7) (Jul-Aug): 390-406.
- Gfeller, K., D. Jiang, J. J. Oleson, V. Driscoll, and J. F. Knutson. 2010. Temporal stability of music perception and appraisal scores of adult cochlear implant recipients. *Journal of the American Academy of Audiology* 21 (1) (Jan): 28-34.
- Gfeller, K., J. Oleson, J. F. Knutson, P. Breheny, V. Driscoll, and C. Olszewski. 2008. Multivariate predictors of music perception and appraisal by adult cochlear implant users. *Journal of the American Academy of Audiology* 19 (2) (Feb): 120-34.
- Gfeller, K., C. Turner, M. Mehr, G. Woodworth, R. Fearn, J. F. Knutson, S. Witt, and J. Stordahl. 2002. Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants International* 3 (1) (Mar): 29-53.
- Gfeller, K., C. Turner, J. Oleson, X. Zhang, B. Gantz, R. Froman, and C. Olszewski. 2007. Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear and Hearing* 28 (3) (Jun): 412-23.
- Goorhuis-Brouwer, SM, and AM Schaerlaekens. 2000. Taalpathologie en taaltherapie bij nederlandssprekende kinderen (handbook language development, language pathology and language therapy in dutch speaking children). *Utrecht, De Tijdstroom*.
- Hilliard, R. E. 2003. The effects of music therapy on the quality and length of life of people diagnosed with terminal cancer. *Journal of Music Therapy* 40 (2) (Summer): 113-37.
- Hinderink, J. B., P. F. Krabbe, and P. Van Den Broek. 2000. Development and application of a health-related quality-of-life instrument for adults with cochlear implants: The nijmegen cochlear implant questionnaire. *Otolaryngology--Head and Neck Surgery : Official Journal of American Academy of Otolaryngology-Head and Neck Surgery* 123 (6) (Dec): 756-65.
- Kong, Y. Y., R. Cruz, J. A. Jones, and F. G. Zeng. 2004. Music perception with temporal cues in acoustic and

electric hearing. *Ear and Hearing* 25 (2) (Apr): 173-85.

- Kong, Y. Y., J. M. Deeks, P. R. Axon, and R. P. Carlyon. 2009. Limits of temporal pitch in cochlear implants. *The Journal of the Acoustical Society of America* 125 (3) (Mar): 1649-57.
- Krabbe, P. F. M., J. B. Hinderink, and P. van den Broek. 2000. The effect of cochlear implant use in postlingually deaf adults. *International Journal of Technology Assessment in Health Care* 16 (03): 864-73.
- Lassaletta, L., A. Castro, M. Bastarrica, R. Perez-Mora, B. Herran, L. Sanz, M. J. de Sarria, and J. Gavilan. 2008. Musical perception and enjoyment in post-lingual patients with cochlear implants. *Acta Otorrinolaringologica Espanola* 59 (5) (May): 228-34.
- Lassaletta, L., A. Castro, M. Bastarrica, R. Perez-Mora, R. Madero, J. De Sarria, and J. Gavilan. 2007. Does music perception have an impact on quality of life following cochlear implantation? *Acta Oto-Laryngologica* 127 (7) (Jul): 682-6.
- Leal, M. C., Y. J. Shin, M. Laborde, M. Calmels, S. Verges, S. Lugardon, S. Andrieu, O. Deguine, and B. Fraysse. 2003. Music perception in adult cochlear implant recipients. *Acta Oto-Laryngologica* 123 (7): 826-35.
- Looi, V., and J. She. 2010. Music perception of cochlear implant users: A questionnaire, and its implications for a music training program. *International Journal of Audiology* 49 (2): 116-28.
- Looi, V., K. Gfeller, and V. Driscoll. 2012. Music appreciation and training for cochlear implant recipients: A review. *Seminars in Hearing* 33 (4) (Nov 1): 307-34.
- Looi, V., H. McDermott, C. McKay, and L. Hickson. 2008. Music perception of cochlear implant users compared with that of hearing aid users. *Ear and Hearing* 29 (3) (Jun): 421-34.
- McDermott, H. J. 2004. Music perception with cochlear implants: A review. *Trends in Amplification* 8 (2): 49-82.
- Migirov, L., J. Kronenberg, and Y. Henkin. 2009. Self-reported listening habits and enjoyment of music among adult cochlear implant recipients. *Annals of Otology, Rhinology & Laryngology* 118 (5): 350-5.
- Mirza, S., SA Douglas, P. Lindsey, T. Hildreth, and M. Hawthorne. 2003. Appreciation of music in adult patients with cochlear implants: A patient questionnaire. *Cochlear Implants International* 4 (2): 85-95.
- Parbery-Clark, A., E. Skoe, and N. Kraus. 2009. Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 29 (45) (Nov 11): 14100-7.
- Parbery-Clark, A., E. Skoe, C. Lam, and N. Kraus. 2009. Musician enhancement for speech-in-noise. *Ear and Hearing* 30 (6) (Dec): 653-61.
- Philips, B., B. Vinck, E. De Vel, L. Maes, W. D'Haenens, H. Keppler, and I. Dhooge. 2012. Characteristics and determinants of music appreciation in adult CI users. *European Archives of Oto-Rhino-Laryngology* 269 (3) (Mar): 813-21.
- Robinson, K., S. Gatehouse, and G. G. Browning. 1996. Measuring patient benefit from otorhinolaryngological surgery and therapy. *Annals of Otology, Rhinology & Laryngology* 105 (6): 415-22.
- Salimpoor, V. N., M. Benovoy, K. Larcher, A. Dagher, and R. J. Zatorre. 2011. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience* 14 (2): 257-62.
- Shannon, R. V., Q. J. Fu, and J. Galvin 3rd. 2004. The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Oto-Laryngologica.Supplementum* (552) (552) (May): 50-4.
- Stickney, G. S., F. G. Zeng, R. Litovsky, and P. Assmann. 2004. Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America* 116 (2) (Aug): 1081-91.
- Walworth, D., C. S. Rumana, J. Nguyen, and J. Jarred. 2008. Effects of live music therapy sessions on quality of life indicators, medications administered and hospital length of stay for patients undergoing elective surgical procedures for brain. *Journal of Music Therapy* 45 (3) (Fall): 349-59.
- Won, J. H., W. R. Drennan, and J. T. Rubinstein. 2007. Spectral-ripple resolution correlates with speech

3

reception in noise in cochlear implant users. *Journal of the Association for Research in Otolaryngology* 8 (3): 384-92.

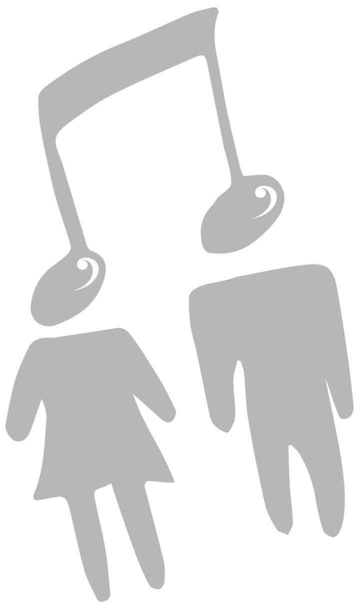
- Won, J. H., W. R. Drennan, R. S. Kang, and J. T. Rubinstein. 2010. Psychoacoustic abilities associated with music perception in cochlear implant users. *Ear and Hearing* 31 (6) (Dec): 796-805.
- Wright, R., and R. M. Uchanski. 2012. Music perception and appraisal: Cochlear implant users and simulated cochlear implant listening. *Journal of the American Academy of Audiology* 23 (5) (May): 350,65; quiz 379.
- Zhao, F., Z. Bai, and D. Stephens. 2008. The relationship between changes in self-rated quality of life after cochlear implantation and changes in individual complaints. *Clinical Otolaryngology* 33 (5): 427-34.

Chapter 4

Gender categorization is abnormal in cochlear-implant users

Christina Fuller [1,2,] Etienne Gaudrain [1,2] Jeane Clarke [1,2]
John J. Galvin III [1,2,3,] Qian-Jie Fu [1,2,3,] Rolien Free [1,2]
and Deniz Başkent [1,2]

JARO, 2014 Dec;15(6):1037-48. Epub 2014 Aug 30



1 Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

2 Graduate School of Medical Sciences (Research School of Behavioural and Cognitive Neurosciences), University of Groningen, The Netherlands.

3 Division of Communication and Auditory Neuroscience, House Research Institute, Los Angeles, CA 957; Department of Head and Neck Surgery, David Geffen School of Medicine, UCLA, Los Angeles, CA 995-1624.

ABSTRACT

In normal hearing (NH), the perception of the gender of a speaker is strongly affected by two anatomically related vocal characteristics: the fundamental frequency (F0), related to vocal pitch, and the vocal tract length (VTL), related to the height of the speaker. Previous studies on gender categorization in cochlear implant (CI) users found that performance was variable, with few CI users performing at the level of NH listeners. Data collected with recorded speech produced by multiple talkers suggests that CI users might rely more on F0 and less on VTL than NH listeners. However, because VTL cannot be accurately estimated from recordings, it is difficult to know how VTL contributes to gender categorization. In the present study, speech was synthesized to systematically vary F0, VTL, or both. Gender categorization was measured in CI users, as well as in NH participants listening to unprocessed (only synthesized) and vocoded (and synthesized) speech. Perceptual weights for F0 and VTL were derived from the performance data. With unprocessed speech, NH listeners used both cues (normalized perceptual weight: F0=3.76, VTL=5.56). With vocoded speech, NH listeners still made use of both cues but less efficiently (normalized perceptual weight: F0=1.68, VTL=0.63). CI users relied almost exclusively on F0 while VTL perception was profoundly impaired (normalized perceptual weight: F0=6.88, VTL=0.59). As a result, CI users' gender categorization was abnormal compared to NH listeners. Future CI signal processing should aim to improve the transmission of both F0 cues and VTL cues, as a normal gender categorization may benefit speech understanding in competing talker situations.

Key words: Cochlear implants, Gender categorization, Fundamental frequency, Vocal tract length, Vocal characteristics

INTRODUCTION

In "cocktail party" listening conditions, normal hearing (NH) listeners use the voice characteristics of different talkers to track and listen to a target talker. The ability to identify the gender of a voice may help to sort out various talkers in a multi-talker environment, especially when two talkers are speaking at the same time. Voice differences across speakers of the same gender can improve intelligibility of the target speech by more than 20 percentage points (Brungart 2001). Voice differences across gender can increase intelligibility by 50 percentage points (Brungart 2001; Festen and Plomp 1990). NH listeners use two anatomically related vocal characteristics to identify the gender of a talker: (i) the fundamental frequency (F0) of the voice, related to perceived vocal pitch and determined by the glottal pulse rate, and (ii) vocal tract length (VTL)[1], mainly related to the height of the speaker (Fitch and Giedd 1999). F0 and VTL have been shown to similarly influence NH listeners' voice gender identification (Skuk and Schweinberger 2013) and concurrent speech perception (Darwin et al. 2003).

Unlike NH listeners, cochlear implant (CI) users do not benefit from differences in the speaker's gender in competing talker situations (Luo et al. 2009; Stickney et al. 2004). This may be partly due to poor representation and/or perception of voice characteristics. Previous studies have shown that CI users' gender categorization performance is highly variable and generally poorer than that of NH listeners (Fu et al. 2004, 2005; Kovačić and Balaban 2009, 2010; Massida et al. 2013; Wilkinson et al. 2013). It was argued in these studies that CI users might rely more on F0 than NH listeners. In Fu et al. (2005), when the F0s of the talkers were overlapping, CI users' gender categorization performance was poorer than that of NH participants listening to sinewave-vocoded stimuli (68 vs. 92 % correct). Subsequently, Kovačić and Balaban (2009) also observed that gender categorization was particularly difficult for CI listeners when the F0 was within the overlap region between the male and female ranges. Recently, Massida et al. (2013) created a continuum between a typical female voice and a typical male voice using a morphing technique. They observed that CI users had shallower psychometric functions than NH listeners and concluded that categorization of ambiguous voices, around the middle point of the continuum, was more difficult for CI users than for NH listeners.

However, the origins of these difficulties are, as yet, unknown. The studies cited above essentially focus on the role of F0, but VTL could also play a crucial role in the categorization of voices, especially when the F0 cue is ambiguous. For instance, although F0 values were estimated and reported in Fu et al. (2005), there was no attempt to estimate talker VTL

1 VTL affects the center frequency of the formants and is sometimes referred to as 'formant dispersion': lengthening the vocal tract by a given factor results in dividing all formant frequencies by that same factor, equivalent to an homothetic translation of the spectral envelope on a log-frequency axis (a detailed explanation can be found in Patterson et al. 2010). One of the main differences between VTL and F0, unlike for glottal pulse rate, F0, and pitch, there are no commonly defined terms to denote the acoustic and perceptual analogs of VTL. In the present study, we therefore used the term VTL to refer to the physical dimension, the apparent acoustic dimension, as well as the perceived quantity related to this anatomical property.

values. This is probably explained by the fact that, unlike F0, it is difficult to estimate VTL from recordings. To date, the best estimators only achieve between 10 and 30 % root-mean-square-error accuracy (Lammert et al. 2013), which is similar to differences between males and females when measured anatomically (15 %, according to Fant 1970). Thus, it is unclear in Fu et al. (2005) and Massida et al. (2013) to what degree VTL cues might have contributed to CI and NH performance. Moreover, although F0 and VTL seem to be the most important cues for gender categorization in NH listeners (Skuk and Schweinberger 2013), other cues also contribute to gender categorization in recordings of real speech, such as breathiness (Holmberg et al. 1988; Van Borsel et al. 2009) or intonation (Fitzsimons et al. 2001). These cues may be used differently by CI users, further complicating the interpretation of past studies based on natural utterances by male and female speakers. One indication that VTL cues might be particularly degraded comes from a study by Mackersie et al. (2011) who observed that listeners with mild to severe hearing loss above 1 kHz could not benefit from VTL differences in a concurrent sentence experiment. By extension, it seems likely that CI listeners might also have difficulties with this cue, but this remains to be shown.

In the present study, we focused on the role of F0 and VTL for gender categorization in NH and CI listeners, by artificially manipulating these two dimensions in stimuli resynthesized from one single female voice. Although the reduced spectral resolution inherent to CI sound transmission notoriously degrades the F0 representation, pitch perception remains possible on the basis of temporal cues (see Moore and Carlyon 2005 for a review). In particular, it can be expected that F0 differences of about one octave that separate typical male from typical female voices would be accessible. However, when the F0 difference is smaller, this cue might become more ambiguous and less useful. VTL, on the other hand, affects the location of the formants (see Fig. 1). In other words, accurate perceptual estimates of VTL rely on accurate perception of the formant peak locations. The limited spectral resolution of the implant, therefore, would be expected to severely hinder the perception of this cue, although such an effect has not been documented. The electrodograms in Fig. 1 suggest that the typical VTL difference between a male and a female voice results in a shift of the electrical stimulation pattern by one electrode. Different spectral resolution measures yield slightly different predictions regarding the detectability of such a shift (see "DISCUSSION" for more details). It could thus also be the case that impaired VTL perception prevents voices with ambiguous F0s from being properly categorized.

The purpose of the present study was to directly measure and characterize the contribution of F0 and VTL cues to gender categorization by CI users as compared to NH listeners. Because VTL cannot be easily estimated from recordings of real speech, speech stimuli were resynthesized to effect systematic manipulation of F0 and apparent VTL cues. Gender categorization with resynthesized speech was measured as a function of VTL and F0 in CI users and in NH subjects listening to non-vocoded and vocoded versions of the synthesized stimuli. Perceptual weights for F0 and VTL were derived from the CI, NH, and NH-vocoded gender categorization data. We predicted that the poor spectral resolution of the implant would affect the relative weights attributed to VTL and F0. A similar prediction was also made for NH listeners tested with degraded spectral cues in the vocoded condition.

METHODS

Participants

Nineteen postlingually deafened CI users (11 male and 8 female, mean age=64.6 years, range=28–78 years) with more than 1 year of CI experience (mean experience=4.6 years, range=1–12 years) were recruited. One CI user was bilaterally implanted. The details of all CI participants are shown in Table 1.

Table 1 – Details of the CI participants.

Subject number	Gender	Years of CI use	Cochlear implant	Speech processor	Rate of stimulation
1	male	9	CI24R CS	CP810	900
2	male	5	HiRes 90K Helix	Harmony	3712
3	male	4	HiRes 90K Helix	Harmony	849
4	male	1	CI24RE CA	CP810	900
5	female	4	HiRes 90K Helix	Harmony	2184
6	female	12	CI24R k	CP810	900
7	male	2	CI24RE CA	CP810	900
8	male	5	CI24RE CA	Freedom	900
9	female	2	CI24RE CA	CP810	900
10	female	3	CI512	CP810	900
11	male	6	HiRes 90K Helix	Harmony	2900
12	male	4	HiRes 90K Helix	Harmony	1740
13	female	3	CI24RE CA	CP810	900
14	male	8	CI24R CA	CP810	900
15	male	5	CI 11+11+2M	Freedom	900
16	female	2	CI24RE H	CP810	900
17	male	2	CI24RE CA	CP810	900
18	female	1	CI24RE CA	CP810	900
19	female	9	CI24R CA	Freedom	900

This study was conducted in parallel with Fuller et al. (2014), where a musician effect was explored on gender categorization, and the same non-musician NH listeners comprised the control group in both studies. The criterion for non-musician was to have not received musical training within the 7 years preceding the study. The motivation for excluding musicians was that it was suspected that musicians might make different use of voice cues than non-musicians, especially in degraded conditions (which was confirmed by Fuller et al. 2014). As such, non-musician NH listeners were thought to be a better control group for CI listeners, who also tend to be not musically involved post-implantation (e.g., Fuller et al. 2012), than

NH listeners with extensive musical expertise. The NH control group of the present study comprised 19 NH participants (3 male and 16 female; mean age=22.1 years, range=19–28 years), who were a subset of the 25 NH non-musician listeners reported in Fuller et al. (2014). NH participants were audiometrically selected to have pure tone thresholds better than 20 dB HL at frequencies between 250 and 4,000 Hz. All participants were native Dutch speakers, with no neurological disorders. The study protocol was approved by the Medical Ethical Committee of the University Medical Center Groningen. Detailed information about the study was provided to the participants before data collection, and written informed consent was obtained. All subjects received financial reimbursement for their participation.

STIMULI

Speech synthesis

The sources for subsequent speech synthesis were four meaningful Dutch words in CVC format (“bus,” “vaak,” “leeg” and “pen,” meaning “bus,” “often,” “empty,” and “pencil,” respectively), taken from the NVA corpus (Bosman and Smoorenburg 1995). The source speech tokens were spoken by a single Dutch female talker. The average word duration was 0.83 s and the average F0 was 201 Hz. The VTL was estimated to be 13.5 cm, based on an average height of 169 cm for Dutch women and the regression between VTL and height reported by Fitch and Giedd (1999).

The source speech tokens were manipulated using the STRAIGHT software (v40.006b; Kawahara et al. 1999), implemented in MATLAB. Both the F0 and the VTL of the source female voice were manipulated to obtain a male voice at the extreme parameter values, where the F0 was decreased by an octave and the VTL was increased by 23 % (resulting in a downward spectral shift of 3.6 semitones). To achieve this in STRAIGHT, the speech signal was first decomposed into the F0 contour and the spectral envelope. All values of the F0 contour were then multiplied by a specific factor, resulting in a change in the average F0 while preserving the relative fluctuations. The VTL lengthening was effected by compressing the extracted spectral envelope toward the low frequencies. The modified components were then recombined via a pitch synchronous overlap-add resynthesis method. In previous studies with similar manipulations, Clarke et al. (2014) confirmed that the chosen F0 and VTL values, applied together, indeed made the listeners perceive a talker of a different gender than the original one, and Fuller et al. (2014) confirmed these values provided a full characterization of gender categorization from the female’s voice to that of a man’s.

In the present study, similar to the studies by Clarke et al. and Fuller et al., intermediate steps were created between the source female voice and the target male talker. The F0 was varied to be 0, 3, 6, 9, or 12 semitones below the F0 of the original female source, which corresponds to changes of 0, 19, 41, 68, and 100 % or average F0 values of 201, 169, 142, 119, and 100 Hz. The VTL was varied to be 0.0, 0.7, 1.6, 2.4, 3.0, or 3.6 semitones, i.e., 0, 4, 7, 14, 19, and 23 % longer than the VTL of the female source, corresponding to lengths of 13.5, 14.1, 14.8, 15.5, 16.1, and 16.6 cm. These combinations produced 30 different voices and resulted in a total of 120 stimuli (5 F0 values×6 VTL values×4 words). All stimuli were resynthesized, even when the original values of F0 and VTL were used. Smith et al. (2007) estimated distributions of natural voices in the F0–VTL plane based on Peterson and Barney (March 1952) and Fitch and Giedd (1999). Using these estimates, we calculated that all the synthesized voices were within 99.7 % of the adult population, and 22 of the 30 voices were within 95 %.

Vocoder processing

Similar to the studies by Fu et al. (2004, 2005), a simple acoustic CI simulation was used in the form of an eight-channel, sinewave vocoder. The vocoder was based on the continuous interleaved sampling strategy (Wilson et al. 1991) and was implemented using the Angelsound™ software (Emily Shannon Fu Foundation, http://www.angelsound.tigerspeech.com/). An eight-channel vocoder was used because it has been shown to yield both gender categorization and speech intelligibility performance similar to that of the best performing CI users (Fu et al. 2004, 2005; Friesen et al. 2001). Both of these are an indication that the eight-band vocoder likely delivers spectral resolution functionally similar to that of better-performing CI users. Despite this functional similarity, it should be noted that this type of vocoder does not accurately reflect the processes happening in actual implants and is here merely used to provide an indication of how degraded spectral cues can affect the task in normal hearing. The input frequency range was 200–7,000 Hz. The acoustic input was bandpass-filtered into eight frequency analysis bands using fourth order Butterworth filters. The band cutoff frequencies were distributed according to the Greenwood (1990) frequency-place formula. For each band, a sinusoidal carrier was generated; the frequency of the sinewave carrier was equal to the center frequency of the analysis filter (i.e., the geometric mean of the band cutoff frequencies). The temporal envelope was extracted from each band using half-wave rectification and lowpass filtering with a Butterworth filter (cutoff frequency=160 Hz, fourth order). These envelopes modulated the corresponding sinusoidal carriers. Finally, the modulated carriers were summed and the overall level was adjusted to be the same level as the original speech token. Figure 1 shows from the left to the right panel the spectra of the generated sounds, the electrodograms, and the total amount of current per channel accumulated over the duration of the vowel, respectively. The middle row shows the stimulus resynthesized in STRAIGHT, with the F0 and VTL of the original female voice. The top row shows the stimulus resynthesized with only the F0 shifted by an octave down. The bottom row shows the stimulus with only the VTL made 23 % longer, which resulted in all formants being shifted down by 3.6 semitones.

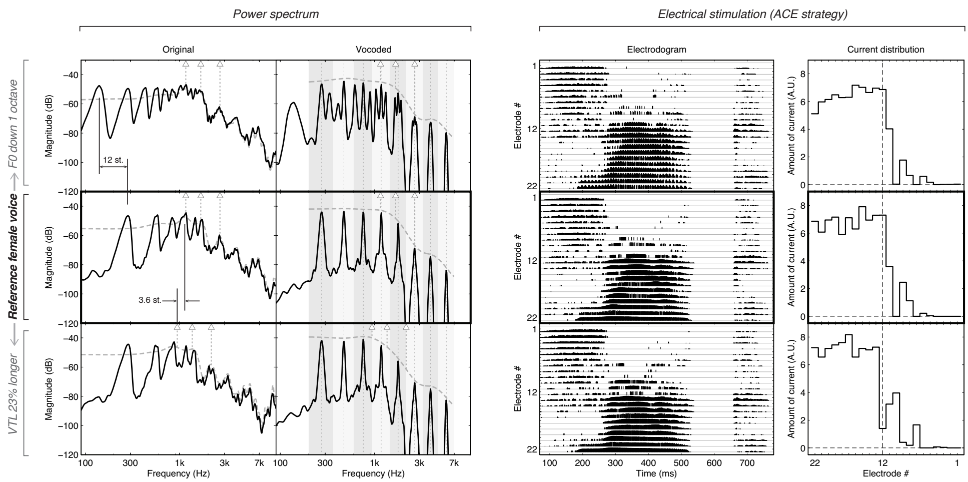


Figure 1 – Power spectrum, waveform and electrodogram of the vowel /aa/ in 'Vaak'. A different voice is represented per row. The stimulus resynthesized with the original parameters of the female voice is shown in the middle row. The *top ro*w shows the F0 changes only, by an octave down. The *bottom row* shows the VTL changed to be made 23% longer, which results in shifting all the formants down by 3.6 semitones (st). The *left panel* shows, over the duration of the vowel, the spectra, for the non-vocoded (*left column*, noted 'Original') and vocoded (*right column*) versions of the stimulus. The spectrum itself is shown by the solid black line, visualizing the harmonics and/or the sinusoidal carriers of the vocoder. The spectral envelope is represented by the dashed gray line as extracted by STRAIGHT for the non-vocoded sounds on the left, and as an interpolation between the carriers for the vocoded sounds on the right. The locations of the first three formants, based on a visual inspection of the envelope, are pointed out by the triangles and stems, for both the left and the right columns. The analysis filter bands of the vocoder are shown in the gray areas in the right column, whereas the sine-wave carrier's frequency is shown with a dotted line. The *right panel* shows the electrical stimulation as obtained with the Nucleus Matlab Toolbox (v4.31, Cochlear Limited, Australia) using an ACE strategy with a default frequency map. The *left column* shows the electrodogram for the whole word, while the *right column* shows the total amount of current per channel accumulated over the duration of the vowel. The vertical line dashed line in this column locates the middle electrode.

PROCEDURE

All synthesized stimuli, with or without vocoding, were presented using Angelsound™ software (Emily Shannon Fu Foundation, http://www.angelsound.tigerspeech.com/). The stimuli were routed via a PC with an Asus Virtuoso Audio Device soundcard (ASUSTeK Computer Inc, Fremont, USA), converted to an analog signal via a DA10 digital-to-analog converter of Lavry Engineering Inc. (Washington, USA), and then played at 65 dB SPL in free field in an anechoic chamber. The participants were seated at a distance of 1 m from the speaker (Tannoy Precision 8D; Tannoy Ltd., North Lanarkshire, UK). During testing, the participant heard a randomly selected stimulus and their task was to select one of two response buttons shown on screen labeled "man" or "vrouw" (i.e. "man" or "woman", in Dutch), to indicate the gender of the talker. The participants replied on an A1 AOD 1908 touch screen (GPEG International, Woolwich, UK). CI users were tested with their own clinical processor. The CI participants were instructed to use their everyday clinical volume and sensitivity settings and to use these settings throughout testing. CI listeners were tested with non-vocoded stimuli. NH listeners were tested first with non-vocoded stimuli and then with vocoded stimuli.

Participant responses were directly scored by the program. NH listeners were not naïve to the vocoding processing as they had participated in similar experiments before. No training was provided to either participant group for the gender recognition task. The gender categorization task lasted for 10 min. This resulted in a total testing time of approximately 20 min for NH participants and 10 min for CI users.

Statistical analysis

All statistical analyses were done in R (version 3.01, R Foundation for Statistical Computing, Vienna, Austria) using the lme4 package (version 1.0-5, Bates et al. 2013). A generalized linear mixed effects model with a logit link function was used following the method described by Jaeger (2008). The model selection started from the full factorial model in lme4 syntax:

```
score ~ f0*vtl*moh + (1+f0*vtl | subject)
```

The variable score is the proportion of “man” responses. The *f0* and *vtl* factors are normalized dimensions defined as $f0 = -\Delta F0/12 - 1/2$ and $vtl = \Delta VTL/3.6 - 1/2$ where ΔF0 and ΔVTL represent the F0 and VTL difference in semitones relative to the original voice. With these normalized dimensions, the point ($f0$=–0.5, vtl=–0.5) represents the original female voice, while the point ($f0$=0.5, vtl=0.5) represents the artificially created male voice. The factor *moh* codes the mode of hearing (NH, NH-vocoded, or CI). The notation “(…|…)” denotes the random effect, here per subject, with “1” thus representing a random intercept per subject. The full factorial model had an Akaike information criterion (AIC)=6342, a Bayesian information criterion (BIC)=6492, and a log-likelihood=−3149. The full factorial model was not significantly different from the simpler model below [$\chi2(7)=13.45$, $p=0.062$], which was then retained as reference:

```
score ~ (f0+vtl)*moh + (1+(f0+vtl) | subject)
```

This model had an AIC=6341, a BIC=6443, and a log-likelihood=−3155. This model has random intercept per subject, as well as random slopes for *f0* and *vtl*, also per subject. Effects for each factor were then tested using the χ2 statistic and *p*-values obtained from the likelihood ratio test comparing the model without the factor of interest against the reference model. In order to compare modes of hearing, the model above was applied to subsets of the data, excluding one mode of hearing at a time and testing the *moh* effect and its interactions within the remaining dataset. Because there were only three comparisons, no correction for multiple comparisons was applied but note that none of the obtained statistics would have changed significance even with a correction as stringent as the Bonferroni correction.

To quantify the contribution of the F0 and VTL, a simpler logistic regression model was used (as described, for instance, by Peng et al. 2009). The "perceptual weights" for each cue were estimated as the coefficients for the *f0* and *vtl* factors in the logistic regression model. In other words, the cue weights are expressed as a and b in the equation logit(*score*) = *a f0* + *b vtl* + ε, where ε is the subject-dependent random intercept. Given the coding of the *f0* and *vtl* variables, the cue weights represent variations in log odd ratios over the entire course of change along each of the cues. Cue weights for groups of subjects are accompanied with their associated Wald statistic z. Individual cue weights were also obtained using the model used for the statistical analyses, i.e., with random *f0* and *vtl* effects. These are reported in Table 2.

RESULTS

In this study, there was no "correct" answer for gender categorization, as all stimuli were resynthesized to be between a woman's voice and a man's voice. Therefore, the categorization judgment of NH group was considered to be the "normal" gender categorization, and CI and NH-vocoded performance were evaluated with respect to this normal performance. Figure 2 shows the results for the three modes of hearing in relation to the normal performance in this test, as is defined by the performance of NH listeners. The normal data are the NH results that are ordered from most strongly judged female voice conditions in the left to most strongly judged male voice conditions in the right. The figure clearly shows a more variable and abnormal pattern for the gender categorization in CI users compared to both the NH and the NH-vocoded modes of hearing. The NH-vocoded mode of hearing also differs from the normal categorization, but there was less variation in their judgment than the real CI users.

Figure 3 shows the average and individual results in more detail, for all conditions tested, and separately for the NH (top), the NH-vocoded (middle), and CI (bottom) modes of hearing. The comparison between the top and bottom panels again shows the discrepancy between NH and CI listeners. With non-vocoded speech (top panel), NH responses gradually shift from female to male as the VTL or F0 are increased. With the vocoded speech (middle panel) or with real CI users (bottom panel), VTL had little effect on gender categorization. Compared to VTL, F0 had a stronger effect on performance both for NH-vocoded group (middle panel) and for real CI users (bottom panel).

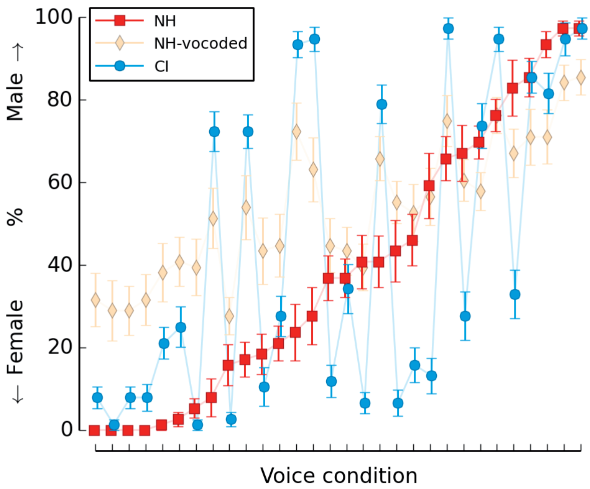


Figure 2– Gender categorization results of NH listeners (red squares), NH listeners tested with vocoded stimuli (CIsim, yellow diamonds), and CI users (blue circles). The x-axis represents the 30 voice conditions ordered according to the NH listeners' average gender categorization, from female on the left, to male on the right. The circles and diamonds show the data for the actual and simulated CI listeners for the same voice conditions. The error bars represent the standard error.

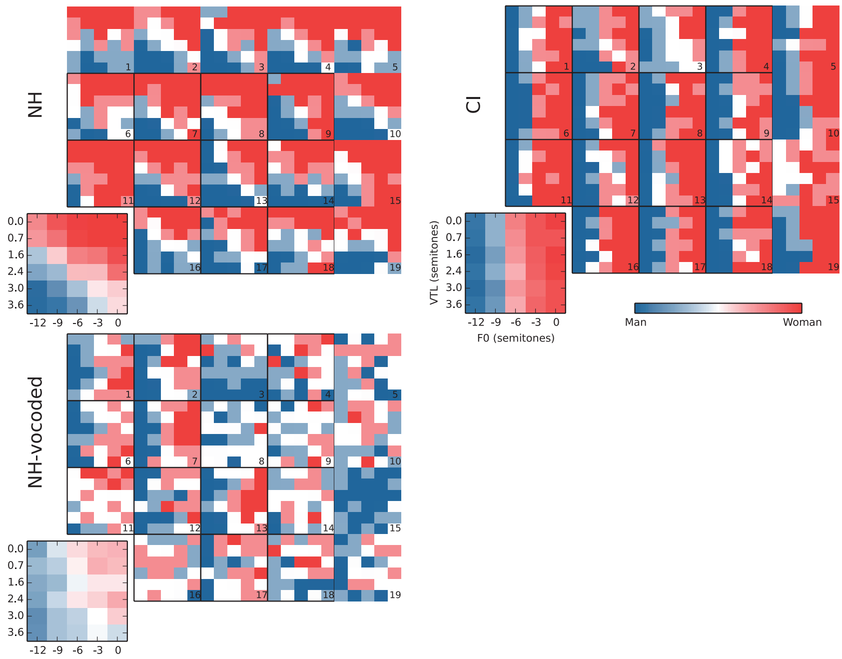


Figure 3 - Individual and average gender categorization judgments, presented as maps in the F0-VTL plane. For each mode of hearing, the smaller panels numbered 1 to 19 show the individual maps where each pixel corresponds to a combination of F0 and VTL, while blue corresponds to 100% "man" responses and red corresponds to 100% "woman" responses.

On average F0 [$\chi^2_{(6)}$=2184, p<0.0001] and VTL [$\chi^2_{(6)}$=958.4, p<0.0001] both had a significant effect on gender categorization and both interacted with the mode of hearing [F0: $\chi^2_{(2)}$=105.3, p<0.0001; VTL: $\chi^2_{(2)}$=420.1, p<0.0001]. Mode of hearing itself also had a main effect on the results [$\chi^2_{(2)}$=271.2, p<0.0001]. These effects are detailed in the following sections, and perceptual weights are reported for each of these cues and modes of hearing.
Individual logistic regression coefficients are reported in Table 2.

Table 2. - Individual logistic regression coefficients for each subject in each mode of hearing. The 'Intercept', 'F0' and 'VTL' columns correspond, respectively, to ε, *a* and *b* coefficients of the regression equation given in the methods section. Summary statistics are given at the bottom of the table. See the section on statistical analyses for details about the calculation of these coefficients. Note that the average of the individual coefficients do not exactly match the coefficients reported in text which result from fitting the logistic regression model to the population (i.e. without F0 and VTL as random effects).

	NH								
	Non-vocoded			Vocoded			CI		
	Intercept	F0	VTL	Intercept	F0	VTL	Intercept	F0	VTL
1	-0.79	1.52	6.17	0.12	3.55	-0.34	-0.75	7.41	0.44
2	-0.43	4.44	5.68	0.09	3.77	1.31	-0.17	5.41	1.19
3	-0.48	5.29	5.29	0.96	0.30	2.03	0.01	3.96	1.33
4	-1.01	3.70	6.13	0.14	1.23	1.44	-1.19	10.33	0.05
5	-1.36	2.50	6.21	0.34	0.44	0.93	-1.08	8.62	0.07
6	-1.04	2.39	5.77	0.25	3.07	-0.08	-0.88	10.03	0.39
7	-1.72	4.45	5.87	-0.19	4.83	0.68	-0.78	7.73	0.42
8	-2.17	4.30	5.75	0.48	0.71	-0.01	-0.66	9.16	0.59
9	-1.29	6.31	5.42	0.14	1.48	0.20	-0.28	6.80	0.93
10	-0.51	3.42	6.07	0.05	0.35	0.78	-0.16	8.28	1.24
11	-3.16	5.36	5.52	-0.33	1.31	1.48	-1.52	9.15	-0.32
12	-1.39	3.92	6.05	0.46	2.68	0.59	-0.08	5.90	1.23
13	0.21	5.55	5.23	0.04	4.38	0.45	-1.04	7.75	0.23
14	-0.50	2.40	6.10	0.17	0.96	0.41	-0.34	5.99	0.88
15	-2.35	6.04	5.36	1.63	0.50	-0.24	-1.01	2.42	0.05
16	-0.30	3.10	5.86	0.02	-0.23	0.23	-1.19	10.33	0.05
17	-0.42	4.33	5.62	0.03	2.48	0.52	-0.49	8.18	0.74
18	-0.98	3.92	5.41	0.20	0.83	1.39	-0.45	7.86	0.83
19	-0.74	2.83	6.07	0.48	1.84	0.54	-0.76	9.09	0.61
Min	-3.16	1.52	5.23	-0.33	-0.23	-0.34	-1.52	2.42	-0.32
Max	0.21	6.31	6.21	1.63	4.84	2.03	0.01	10.33	1.33
Mean	-1.08	3.99	5.77	0.27	1.82	0.65	-0.68	7.60	0.58
Std. dev.	0.82	1.34	0.33	0.43	1.51	0.65	0.44	2.12	0.48

Comparisons of modes of hearing

NH listeners (top panel of Fig. 3) gave high weights both to F0 (3.76, z=18.1) and VTL (5.56, z=22.6), indicating that they used both dimensions to estimate the gender of the voices. For NH subjects to completely perceive the female voice as male, both F0 and VTL needed to be changed; changing F0 alone or VTL alone produced less reliable categorization in most cases. In particular, a change of −12 semitones in F0 with no change of VTL produced a male judgment only in 10 % of the trials, illustrating the importance of VTL for gender categorization. Individual weights for VTL (see Table 2) were also remarkably similar across participants (ranging from 5.23 to 6.21, s.d. 0.33) while those for F0 showed larger variability (1.52 to 6.31, s.d. 1.34).

In contrast, CI listeners (bottom panel of Fig. 3) relied more on F0 (6.88, z=25.4) than the NH listeners [$\chi^2_{(1)}$=94.51, p<0.0001] and less on VTL (0.59) than the NH listeners [$\chi^2_{(1)}$=301.2, p<0.0001]. The CI listeners showed a somewhat larger variability across listeners in their sensitivity to both F0 (weights ranging from 2.42 to 10.33, s.d. 2.12) and VTL (weights ranging from −0.32 to 1.33, s.d. 0.48). There was no main effect of mode of hearing between these two groups [$\chi^2_{(1)}$=2.87, p=0.0888] indicating that mode of hearing did not bias gender categorization toward one sex or the other.

4

In the NH-vocoded condition (middle panel of Fig.3), the weights were reduced both for F0 [weight: 1.68; vs. NH: $\chi^2_{(1)}$=66.70, p<0.0001] and VTL [weight: 0.63; vs. NH: $\chi^2_{(1)}$=382.2, p<0.0001]. These perceptual weights obtained for F0 were also different from the one obtained for actual CI listeners [$\chi^2_{(1)}$=404.8, p<0.0001], but those obtained for VTL were not significantly different [$\chi^2_{(1)}$=0.034, p=0.85]. Finally, in the NH-vocoded condition, listeners showed large inter-individual variability: weights for F0 ranged from -0.23 to 4.84 (s.d. 1.51), and weights for VTL ranged from 0.34 to 2.03 (s.d. 0.65).

Within group factors for the CI listeners

Although the variability across CI listeners was relatively small, a number of factors were tested for significance by adding them to the reference model. We found that the type of *speech processor* of the implant had a significant main effect on gender categorization [$\chi^2_{(2)}$=12.929, p=0.0016], but this effect did not interact with either F0 or VTL. The Freedom and CP810 processors from Cochlear Limited (Australia) were not different from each other [p=0.84], but the users of the Harmony processor from Advanced Bionics AG (Switzerland) were significantly more likely to answer 'female' than the other participants [p<0.0001]. This could be a confound with the effect of *rate of stimulation* [$\chi^2_{(1)}$=6.893, p=0.0087], which also did not interact with F0 and VTL: overall, participants with higher stimulation rates (i.e. using the Harmony processor) had a higher tendency to answer 'female' than those with lower rates. This effect was not significant anymore when the effect of processor was partialled out.

Another factor that could potentially influence gender categorization is the *type of electrode array* of the implant. Some arrays are designed to place electrodes closer to the modiolus and limit cochlear damage during insertion. In our group of subject this might be the case for users of 'CI24R CS' and 'CI24RE H'. However, only two of the 19 CI participants had electrode arrays that differed from the others, and inspection of the individual regression coefficients for these participants did not reveal a particular pattern.

Further examining individual results, it appears that four participants had perceptual weights greater than 1.0 for VTL (subject number 2, 3, 10 and 12). Looking at the history, device, duration of implantation, age or gender of these participants, however, we could not find a common trait. Similarly, the four listeners who had the highest perceptual weights for F0 had nothing in common: they used different devices, had different ages and were of different sex.

Finally, two of the participants used the Fidelity 120 strategy of Advanced Bionics. This strategy involves current steering and thus offers the possibility to deliver peaks of the spectrum at their exact location, which could provide a significant advantage for VTL perception. However, these two listeners showed amongst the smallest perceptual weights for VTL.

Measures of sensitivity

To perform the gender categorization task, the listeners integrate the manipulated cues F0 and VTL (in addition to other non-manipulated cues) into a single judgment. This process yields data that can be represented in a three-dimensional space with F0, VTL, and gender categorization as the three dimensions (as displayed in Fig. 3). For each participant, the two perceptual weights, resulting from the cue weighting analysis, define a plane in the logit F0–VTL space. The slope of this surface represents the sensitivity in perceiving the gender difference in stimuli. The maximal slope, or the score gradient, represents the absolute sensitivity independent of the cue that is used and can be calculated as where a and b are the coefficients for *f0* and *vtl* as defined in the logistic regression. Another slope can be calculated along the straight line between the male and the female voice. This diagonal is similar to the line followed by the continuum of voices used in Massida et al. (2013). The slope along this line, calculated as , thus reflects the sensitivity in a way that is comparable to that of Massida et al. (2013). Note that none of these slopes give any indication about the normal behavior by themselves, and they only bear information about how sensitive participants are to any of the cues used in a specific task.

The values for smax and sdiag were calculated for each participant and compared across groups. We found that maximal slopes smax were similar for NH (7.12, s.d. 0.56) and CI (7.65, s.d. 2.09) listeners [t(20.6)=1.06, p=0.29]. However, when comparing slopes along the diagonal, CI users (5.78, s.d. 1.39) did show lower slopes than NH listeners [6.90, s.d. 0.77; t(28.07)=−3.07, p=0.0048].

DISCUSSION

In this study, gender categorization by CI users was shown to be abnormal relative to NH performance with unprocessed speech. By systematically varying F0 and VTL cues with synthesized stimuli, we found that CI users' gender categorization mainly depends on F0 cues, with nearly no contribution of VTL cues. This is an important finding, as F0 alone or VTL alone is not sufficient for the normal categorization of gender.

Normal Gender Categorization

In this study, "normal" gender categorization was defined as NH performance with non-vocoded speech. These results are in accordance with data previously reported in literature that also showed NH subjects to rely equally strongly on both F0 and VTL cues for gender categorization (Skuk and Schweinberger 2013; Smith and Patterson 2005; Smith et al. 2007). Only when both VTL and F0 were changed was the source female voice completely perceived as male. When the source female VTL was retained, even the largest F0 change (−12 semitones) only resulted in a "male" judgment in less than 10 % of the trials. Reciprocally, when the source female F0 was retained and only VTL was changed (by 3.6 semitones), the voice was judged as "male" only in about 30 % of the trials. These results are comparable to those obtained in previous gender categorization studies (Smith and Patterson 2005; Smith et al. 2007) and emphasize the importance of both vocal characteristics.

Gender Categorization by CI Listeners

CI gender categorization was abnormal relative to NH performance with unprocessed speech. Different from NH performance, CI users' weighted F0 cues very strongly and VTL cues almost not at all in the categorization. These results therefore bring strong evidence to what was indirectly suggested in previous studies, namely, that CI users primarily rely on F0 cues for gender categorization (Fu et al. 2004, 2005; Kovačić and Balaban 2009, 2010). However, further, the present results also showed that overreliance on F0 cues may cause CI users to make abnormal judgments of a talker's gender.

Unlike for the NH listeners, the voice presented in the experiment never seemed to be ambiguous to the CI participants. For NH listeners, 7 of the 30 voices produced average male judgments between 35 and 65 %. For the CI listeners, none of the voices produced a judgment in that range. This is in apparent contrast with the results of Massida et al. (2013) who reported that the gender categorization deficit in CI compared to NH listeners was "stronger for ambiguous stimuli" in the continuum between a male and a female voice. This conclusion was supported by the fact that the psychometric functions for their CI participants were 58 % shallower than for their NH participants. In our study, instead of using a unidimensional continuum, we measured gender categorization on a bidimensional space. Sensitivity in such a space is captured by the maximal slope of the two-dimensional psychometric function, i.e., the norm of the gradient of the plane fitted to the logit scores as described in the last part of the "RESULTS" section. With this sensitivity measure, we found that CI listeners showed at least as high sensitivity as NH listeners on average. In other words, the psychometric functions were equally steep for CI and NH listeners, but their orientation in the F0–VTL plane was different. However, when measuring sensitivity along a unidimensional continuum between our female and male voices similar to the one used by Massida et al. (2013), we found results consistent with their findings: that sensitivity along that continuum was smaller for CI listeners than for NH listeners. Our results now bring further explanation that this weaker sensitivity to voice gender is due to a deficit in VTL perception. It is perhaps surprising that CI listeners showed such a strong reliance on F0 cues when pitch perception has been repeatedly reported as defective, or at best, weak, with an implant (see Moore and Carlyon 2005 for a review). However, it is worth noting that the F0 difference separating our male and female voices—one octave—is extremely large compared to F0 difference limens in NH listeners (e.g., Rogers et al. 2006, report F0 difference limens in words of about half a semitone) or even in CI listeners (3.4 semitones, reported in that same study). In other words, while F0 perception is indeed degraded in CI listeners, it remains sufficiently robust to discriminate the pitch of a male voice from that of a female voice.

VTL, on the other hand, could be expected to be more clearly perceived in CIs, as changes along this dimension do not affect the spectral fine structure but the spectral envelope, which is better preserved in the implant. The right-most column of Figure 1 shows electrical stimulation patterns for the voice with the unmodified VTL and the elongated VTL of the male voice. Frequency channels in CIs are typically separated by 2.5 to 3.0 semitones. The VTL separation between the male and female voice, 3.6 semitones, thus results in a shift of the stimulation pattern along the electrode array of about one electrode (Fig. 1, right-most column). Using stimulation patterns comprising one to eight adjacent electrodes (the latter is relatively similar to the stimulation pattern of the vowels in our experiment), Laneau and Wouters (2004) found that CI listeners have just-noticeable differences for place shifts of about 0.5 electrodes. Yet, the CI users in our experiment did not use the VTL cue for gender categorization. Another measure of spectral resolution uses broadband spectral ripple discrimination, where listeners have to discriminate between a spectral ripple pattern and its inverse-phase counterpart. With this method, Anderson et al. (2011) showed that, on average, CI listeners could discriminate phase-inverted spectral ripples up to 1.68 ripple/ octave. The detection of the 3.6-semitone shift in our experiment would require discrimination of 1.67 ripple/octave, so average CI listeners could perhaps just detect this VTL shift. However, on a larger population of CI users, Won et al. (2007) observed that only about 35 % of their participants had discrimination thresholds above 1.44 ripple/ octave. Therefore, it remains unclear whether the VTL shift could be detected at all by the

CI listeners. From these considerations, two hypotheses can thus be formulated. The first one is that although the difference of VTL is visible on the electrodogram, the wide spread of excitation of electrical stimulation prevents this cue from being available in the neural activity pattern. In other words, the effective spectral resolution of electrical stimulation is not sufficient for this cue to be perceived. A direct way to test this hypothesis would be to measure VTL difference limens in CI listeners. The second hypothesis is that this cue remains available to some extent in the neural representation but is either too weak or too distorted to be reliably used for gender categorization. The place–frequency mismatch that results from the fact that electrode arrays cannot be inserted all the way to the apex, for instance, could distort (without removing) the representation of this cue, as previously suggested by Kovačić and Balaban (2009, 2010). In such a context, CI listeners would overly rely on the more robust cue that is available, i.e., pitch. If this hypothesis was verified, i.e., the VTL cue was only distorted but not entirely destroyed, specific training could improve its usability.

Gender Categorization with Vocoded Stimuli

Compared to NH performance with non-vocoded speech, the NH-vocoded performance was much poorer, close to 50 % “man”/“woman” responses at all F0–VTL combinations. Such a pattern can be interpreted as increased uncertainty in the responses or lack of agreement across participants. Examination of the logistic regression coefficients showed that F0 and VTL were used less efficiently than in the non-vocoded condition. This is expected since the sinewave vocoder weakened both F0 and VTL cues, compared to unprocessed speech.

However, performance in the NH-vocoded condition was markedly different from real CI users’ performance, suggesting that sinewave vocoding might be too simple a simulation for gender categorization tasks. A notable difference between actual and simulated CI hearing is that, for conditions where the F0 was below 160 Hz, the sinewave vocoder provided not only temporal but also spectral F0 cues to the NH listeners, which are not available to actual CI users. Nevertheless, NH participants did not seem to make a strong use of these F0 cues as the results below and above F0=160 Hz are not markedly different. More importantly, even when F0 cues were present (below 160 Hz), these cues were weaker than in the non-vocoded condition. Because the same NH subjects did the task first with non-vocoded stimuli and then with the vocoded set, they were aware that the voice cues were weaker in the vocoded case relative to the non-vocoded condition, and this could have, in turn, resulted in them relying less on these cues.

Regarding VTL, as the carrier center frequencies of the vocoder were separated by 7.5 semitones on average (or 2.7 mm in cochlear distance, according to Greenwood 1990), VTL differences as small as 3.6 semitones were not expected to be detectable in the vocoded stimuli. Yet, the cue weight for VTL was larger in the NH-vocoded condition than for CI users. This suggests that CI users’ functional spectral resolution was probably poorer than that achieved by the eight independent frequency channels of the vocoder. The specific role of channel interaction in CIs could be investigated in NH listeners using a more elaborate vocoder (e.g., Churchill et al. 2014).

CONCLUSION

The main finding of our study is that CI users have an abnormal gender categorization compared to NH listeners. CI users strongly and almost exclusively use the F0 cue, while NH listeners use both vocal characteristics, F0 and VTL, for gender categorization. This can have practical consequences on everyday situations for CI users as, for a given voice, they may judge gender differently than what it should be. Further, this could also mean that CI users may not be able to use VTL differences to segregate competing talkers, thus contributing to difficulties understanding speech in multi-talker environments. Consequently, although the CI users achieve some gender categorization, as was also shown previously, the present study emphasizes that their ability to do so is not complete and must be considered impaired.

At this point, it remains unclear whether the observed deficiency in VTL perception is because VTL differences are not transmitted by the CI to the auditory nerve (e.g., because of spread of excitation and channel interaction) or, alternatively, whether they are actually transmitted and detected but not reliable enough for accurate gender categorization. Further research is therefore needed to explore whether VTL differences can be detected at all or whether they are simply not interpreted as talker-size differences. Based on such knowledge, appropriate coding schemes or better fitting algorithms for CIs can be developed and abnormal judgment of gender identification can perhaps be corrected.

Another point that will require further investigation is the extent to which other cues may contribute to gender categorization. Although F0 and VTL seem to be the most important factors for gender categorization in NH listeners (Skuk and Schweinberger 2013), other cues such as breathiness (Holmberg et al. 1988; Van Borsel et al. 2009) or intonation (Fitzsimons et al. 2001) could play a more important role in CI listeners.

Finally, the protocol used in the present study was a quick test (10 min only) that characterized how CI users' gender categorization deviates from normal and what specific vocal cues are underutilized. Using such a quick test, new coding strategies or fitting algorithms can be improved to achieve a normal gender categorization, which will likely indicate that vocal characteristics are fully utilized. Because gender categorization and specifically F0 and VTL differences have been shown to facilitate concurrent speech perception, improving their representation in the implant could, in turn, lead to improved speech-in-noise perception by CI users.

Acknowledgements

We would like to thank the participants in this study. Furthermore, we would like to thank

Joeri Smit and Karin van der Velde for their help with collecting the data, as well as Anita Wagner for her advice regarding statistical methods. The fourth author is supported by a NIH R01- DC004792 grant. The sixth author is supported by an otological/neurotological stipendium from the Heinsius-Houbolt Foundation. The last author is supported by a Rosalind Franklin Fellowship from the University Medical Center Groningen, University of Groningen and the VIDI grant 016.096.397 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw). The study is part of the research program of our department: Healthy Aging and Communication.

REFERENCES

- Anderson, E. S., D.A. Nelson, H. Kreft, P.B. Nelson, and A.J. Oxenham. 2011. Comparing spatial tuning curves, spectral ripple resolution, and speech perception in cochlear implant users. *The Journal of the Acoustical Society of America* 130:364–375.
- Bates, D., M. Maechler, B. Bolker, S. Walker. 2013. lme4: Linearmixed-effects models using Eigen and S4.
- Bosman, A. J., and G. F. Smoorenburg. 1995. Intelligibility of dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology* 34 (5): 260-84.
- Brungart, D. S. 2001. Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America* 109 (3): 1101-9.
- Churchill, T. H., A. Kan, M. J. Goupell, A. Ihlefeld, and R. Y. Litovsky. 2014. Speech perception in noise with a harmonic complex excited vocoder. *Journal of the Association for Research in Otolaryngology* 15 (2): 265-78.
- Clarke, J., E.Gaudrain, M. Chatterjee, and D. Başkent. 2014. T'ain't the way you say it, it's what you say–Perceptual continuity of voice and top–down restoration of speech. *Hearing Research* 315 : 80-7.
- Darwin, C. J., D. S. Brungart, and B. D. Simpson. 2003. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America* 114 (5): 2913-22.
- Fant, G. 1971. *Acoustic theory of speech production: With calculations based on X-ray studies of russian articulations*. Vol. 2Walter de Gruyter.
- Festen, J. M., and R. Plomp. 1990. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America* 88 (4): 1725-36.
- Fitch, W. Tecumseh, and J. Giedd. 1999. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America* 106 (3): 1511-22.
- Fitzsimons, M., N. Sheahan, and H. Staunton. 2001. Gender and the integration of acoustic dimensions of prosody: Implications for clinical studies. *Brain and Language* 78 (1): 94-108.
- Friesen, L. M., R. V. Shannon, D. Başkent, and X. Wang. 2001. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America* 110 (2) (Aug): 1150-63.
- Fu, Q. J., S. Chinchilla, and J. J. Galvin. 2004. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology* 5 (3): 253-60.
- Fu, Q. J., S. Chinchilla, G. Nogaki, and J. J. Galvin III. 2005. Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *The Journal of the Acoustical Society of America* 118 (3): 1711-8.
- Fuller, C. D., R.H. Free, B. Maat, D. Başkent. 2012. Musical background not associated with self-perceived hearing performance or speech perception in postlingual cochlear-implant users. *The Journal of the Acoustical Society of America* 132:1009–1016.
- Fuller, C. D., J. J. Galvin 3rd, B. Maat, R. H. Free, and D. Başkent. 2014. The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience* 8 (Jun 30): 179.
- Greenwood, D. D. 1990. A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America* 87 (6): 2592-605.
- Holmberg, E. B., R. E. Hillman, and J. S. Perkell. 1988. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *The Journal of the Acoustical Society of America* 84 (2): 511-29.
- Jaeger, T.F. 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J Mem Lang* 59:434–446.
- Kawahara, H., I. Masuda-Katsuse, and A. de Cheveigne. 1999. Restructuring speech representations using a

pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication* 27 (3): 187-207.

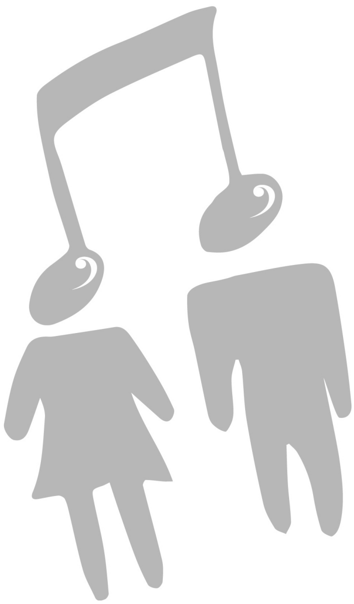
- Kovačić, D., and E. Balaban. 2009. Voice gender perception by cochlear implantees. *The Journal of the Acoustical Society of America* 126 (2): 762-75.
- Kovačić, D., and E. Balaban. 2010. Hearing history influences voice gender perceptual performance in cochlear implant users. *Ear and Hearing* 31 (6) (Dec): 806-14.
- Lammert, A., M. Proctor, and S. Narayanan. 2013. Morphological variation in the adult hard palate and posterior pharyngeal wall. *Journal of Speech, Language, and Hearing Research* 56 (2): 521-30.
- Laneau J.,J. Wouters. 2004. Multichannel place pitch sensitivity in cochlear implant recipients. J Assoc Res Otolaryngol 5:285–294.
- Luo, X., Q. J. Fu, H. P. Wu, and C. J. Hsu. 2009. Concurrent-vowel and tone recognition by mandarin-speaking cochlear implant users. *Hearing Research* 256 (1): 75-84.
- Mackersie, C. L., J. Dewey, and L. A. Guthrie. 2011. Effects of fundamental frequency and vocal-tract length cues on sentence segregation by listeners with hearing loss. *The Journal of the Acoustical Society of America* 130 (2): 1006-19.
- Massida, Z., M. Marx, P. Belin, C. James, B. Fraysse, P. Barone, and O. Deguine. 2013. Gender categorization in cochlear implant users. *Journal of Speech, Language, and Hearing Research* 56 (5): 1389-401.
- Moore, B. C. J., and R. P. Carlyon. 2005. Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In *Pitch.*, 234-277 Springer.
- Patterson, R. D., E. Gaudrain, and T. C. Walters. 2010. The perception of family and register in musical notes. In: Jones MR, Fay RR, Popper AN (eds) Music perception, 1st Edition. Springer, pp13–50
- Peng, S. C., N. Lu, and M. Chatterjee. 2009. Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiol Neurotol* 14:327–337.
- Peterson, Gordon E., and Harold L. Barney. 1952. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America* 24 (2): 175-84.
- Rogers, C.F., E.W. Healy, and A. A. Montgomery. 2006. Sensitivity to isolated and concurrent intensity and fundamental frequency increments by cochlear implant users under natural listening conditions. *The Journal of the Acoustical Society of America* 119:2276–2287
- Skuk, V. G., and S. R. Schweinberger. 2014. Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language, and Hearing Research* 57 (1): 285-96.
- Stickney, G. S., F. G. Zeng, R. Litovsky, and P. Assmann. 2004. Cochlear implant speech recognition with speech maskers. *The Journal of the Acoustical Society of America* 116 (2) (Aug): 1081-91.
- Van Borsel, J., J. Janssens, and M. De Bodt. 2009. Breathiness as a feminine voice characteristic: A perceptual approach. *Journal of Voice* 23 (3): 291-4.
- Wilkinson, E. P., O. Abdel-Hamid, J. J. Galvin, H. Jiang, and Q. J. Fu. 2013. Voice conversion in cochlear implantation. *The Laryngoscope* 123 (S3): S29-43.
- Wilson, B. S., C. C. Finley, D.T. Lawson et al. 1991. Better speech recognition with cochlear implants. Nature 352:236–238.
- Won, J. H., W.R. Drennan, J. T. Rubinstein. 2007. Spectral-Ripple Resolution Correlates With Speech Reception In Noise In Cochlear Implant Users. J Assoc Res Otolaryngol 8:384–392.

Chapter 5

Normal-Hearing Listeners' and Cochlear-Implant Users' Perception of Pitch Cues in Emotional Speech

Steven Gilbers[1], Christina Fuller[2], Dicky Gilbers[1], Mirjam Broersma[3], Martijn Goudbeek[4], Rolien Free[2,3], Deniz Başkent[2,3]

i-Perception, 6-5. (2015)



1 Center for Language and Cognition Groningen, Department of Linguistics, University of Groningen, Groningen, the Netherlands

2 Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, Groningen, the Netherlands

3 Max Planck Institute for Psycholinguistics, the Netherlands/Centre for Language Studies, Radboud University, Nijmegen, the Netherlands

4 Department of Communication and Information Sciences, University of Tilburg, Tilburg, the Netherlands

ABSTRACT

In cochlear-implants, acoustic speech cues, especially for pitch, are delivered in a degraded form. This study's aim is to assess whether due to degraded pitch cues, normal-hearing listeners and cochlear-implant users employ different perceptual strategies to recognize vocal emotions, and, if so, how these differ. Voice actors were recorded pronouncing a nonce word in four different emotions: anger, sadness, joy, and relief. These recordings' pitch cues were phonetically analyzed. The recordings were used to test 20 normal-hearing listeners' and 20 cochlear-implant users' emotion recognition. In congruence with previous studies, high arousal emotions had a higher mean pitch, wider pitch range, and more dominant pitches than low arousal emotions. Regarding pitch, speakers did not differentiate emotions based on valence but on arousal. Normal-hearing listeners outperformed cochlear-implant users in emotion recognition, even when presented with cochlear-implant simulated stimuli. However, only normal-hearing listeners recognized one particular actor's emotions worse than the other actors'. The groups behaved differently when presented with similar input, showing they had to employ differing strategies. Considering the respective speaker's deviating pronunciation, it appears that for normal-hearing listeners, mean pitch is a more salient cue than pitch range, whereas cochlear-implant users are biased towards pitch range cues.

Keywords: acoustic emotion cues, emotion recognition, cue ranking, cochlear-implant, force of articulation

INTRODUCTION

In everyday situations, speech not only conveys a message through semantic content, but also through indexical cues, such as the talker's emotional state. The identification of these indexical cues from acoustic stimuli is essential for robust communication in social situations. However, due to the reduced temporal and spectral speech cues in cochlear-implants (CIs), the prosthetic hearing devices for sensorineural hearing impaired persons, the users of these devices likely do not make full use of these indexical cues. As a consequence, CI users miss out on an important portion of speech communication, which is perhaps a factor contributing to the difficulties CI users encounter in communicating in noisy environments (Fu, Shannon, and Wang 1998; Friesen et al. 2001; Fu and Nogaki 2005).

Former studies showed that even in situations without background noise, adult CI users have difficulties recognizing emotions in speech. Adult CI users were shown to recognize emotions in spoken sentences at an accuracy level ranging from 45% to 51% correct only (House 1994; Pereira 2000; Xin, Fu, and Galvin 2007), in contrast to the high accuracy level of 84% to 90% correct in normal-hearing (NH) listeners (House 1994; Xin, Fu, and Galvin 2007). Luo et al. also showed that emotion recognition was better in NH listeners listening to acoustic simulations of CIs (4-8 channels) than in actual CI users. Moreover, these studies suggested that, due to the aforementioned limitations in temporal and spectral cues in CIs, emotion recognition in CI users is mostly based on the acoustic cues of intensity and duration, but not on the cues of pitch or other voice characteristics. Indeed, the representation of the fundamental frequency (F_0) in CIs – and, therefore, pitch perception in CI users – is notoriously degraded (see (Moore and Carlyon 2005) for a review, as well as (Gaudrain and Baskent 2015) for a discussion on just noticeable differences for voice pitch in CI users and acoustic simulations of CIs). The reduced spectral resolution of the implant is not sufficient to deliver harmonics (in the range of F_0 found in human voices), and therefore F_0 is generally not perceived strongly through spectral cues. However, as the signal delivered in each electrode is modulated by the speech envelope that carries temporal F_0 cues, pitch perception remains limitedly possible. Studies on gender categorization, another task that relies on the perception of temporal and spectral cues of a speaker's voice, confirmed that CI users mostly rely on temporal voice pitch cues, whereas NH listeners can utilize both spectral and temporal voice pitch cues (Fu, Chinchilla, and Galvin 2004; Fu et al. 2005; Kovačić and Balaban 2009; Kovacic and Balaban 2010; Wilkinson et al. 2013; Fuller et al. 2014c).

Recently, Massida et al. (2011) pointed at more central factors, such as a cross-modal reorganization of the speech and voice-related areas of the brain that could also affect perception of indexical cues in CI users, in addition to device-imposed limitations (Massida et al. 2011). Furthermore, auditory deprivation and subsequent CI use can play an important (negative) role in cognitive processing of perceived speech (Ponton et al. 2000). It has been

5

suggested that acoustically impaired listeners may adapt their perceptual strategies by changing the relative importance of acoustic cues in the perceived speech signal (Winn, Chatterjee, and Idsardi 2011; Francis, Baldwin, and Nusbaum 2000; Francis, Kaganovich, and Driscoll-Huber 2008; Fuller et al. 2014c). In other words, the relative importance listeners subconsciously attach to acoustic cues of the perceived speech signal could be determined by the quality of this signal and by which acoustic cues were deemed as more reliable by the listener. Therefore, while pitch perception in CI users has already been shown to be limited and to play a role in reduced emotion recognition in speech, the question still remains whether other factors, such as different processing of the reduced cues to achieve the task, may also play a role.

In this paper, we propose an approach to shed further light on this question. More specifically, we propose a method of assessing relative orderings of acoustic emotion cues in terms of salience by ordering them in a way that is reminiscent of the differences in cue weighting for the recognition of phonemes across languages (see e.g. (Broersma 2005; Broersma 2010; Fitch et al. 1980; Sinnott and Saporita 2000) and rankings in Optimality Theory (Prince and Smolensky 2002). These cues are part of the Force of Articulation Model (Gilbers et al. 2013; van der Scheer, Jonkers, and Gilbers 2014), which encompasses a wide array of both stereotypical phonetic characteristics of high arousal speech (e.g. higher pitch and wider pitch range) and more subtle indicators of force of articulation (e.g. number of dominant pitches in a pitch histogram). The advancement this approach brings to the field is that it allows identification of different listener groups' different biases in auditory perception.

Emotions in speech can be characterized along two dimensions: Valence and Arousal (Russel & Mehrabian, 1977; Russel, 1980; Fontaine, Scherer, Roesch, & Ellsworth, 2007; Goudbeek & Scherer, 2010). The former concerns the difference between positive (e.g. 'joy') and negative emotions (e.g. 'sadness'), and the latter concerns the difference between high arousal (e.g. 'anger') and low arousal emotions (e.g. 'relief') (see Table 1).

Table 1. The selected emotions divided along the Valence and Arousal parameters.

		Valence	
		Positive	*Negative*
Arousal	*High*	Joy	Anger
	Low	Relief	Sadness

While Luo et al. (2007) did not assess Valence and Arousal, a reinterpretation of their results suggests that with respect to mean pitch and pitch range, speakers only differentiate emotions in their speech along the Arousal parameter. In the present study, we aim to replicate this finding by investigating Valence and Arousal more directly; to that end we use the four emotions depicted in Table 1, chosen such that Valence and Arousal are fully crossed. Further, we aim to extend this question to a third pitch parameter, namely the number of dominant pitches, as identified by the PRAAT-analyses. Based on the findings of Luo et al., we expect that regarding pitch-related force of articulation parameters, speakers only differentiate between emotions along the Arousal parameter and not along the Valence parameter. This expectation is supported by studies on another pitch-related force of articulation characteristic, namely the number of dominant pitches in a pitch histogram, which showed that speech often contains multiple dominant pitches in high arousal conditions, whereas speech in low arousal conditions often contains only one dominant pitch (Cook 2002; Cook, Fujisawa, and Takami 2004; Schreuder, van Eerten, and Gilbers 2006; Liberman 2006; Gilbers and Van Eerten 2010).

In this study, in order to investigate whether speakers indeed distinguish between emotions along the Arousal parameter, three pitch-related force of articulation parameters – namely mean pitch, pitch range, and number of dominant pitches – will be acoustically analyzed. Moreover, this study aims to assess which pitch cues are most salient to NH listeners and which ones to CI users. To that end, the aforementioned pitch analyses will also be used to ascertain how individual speakers differ from each other in their production of vocal emotions in nonce words in terms of the degree to which they distinguish between emotions using these pitch cues. Furthermore, this study also assesses listeners' perception of those cues related to production of the vocal emotions in an emotion recognition experiment. By combining the results of the pitch analyses with the emotion recognition data, we will assess which pitch cues are most salient to NH listeners and which ones to CI users.

In sum, the present study focuses on the production of acoustic emotion cues in speech in a nonce word phrase and on the perception of those cues by NH listeners and CI users. Its main aim is to assess if NH listeners and CI users employ different perceptual strategies to recognize vocal emotions, given that the acoustic cues they can use are not the same, and, if so, how their strategies differ. To this end, an approach to map the two groups' perceptual strategies for emotion recognition is proposed. This approach builds on Optimality Theory principles and focuses on different acoustic characteristics of force of articulation. Information on individual speakers' production of pitch-related acoustic emotion cues is combined with information on NH listeners' recognition patterns across speakers – both for normal sound and CI simulated sound – and CI users' recognition patterns across speakers in order to map the two groups' perceptual biases involved in emotion recognition.

METHODS

Participants

Twenty NH listeners (17 females, 3 males; ages 19-35 yr, M = 22.85, SD = 3.87) and 20 post-lingually deafened CI users (9 females, 11 males; ages 28-78 yr, M = 65, SD = 10.86) with more than one year of CI experience participated in the present study. To have a population that represents typical CI users, participants were neither selected based on their device model or their performance with their device, nor controlled for age. All participants were native Dutch speakers with no neurological disorders. All NH listeners had pure tone hearing thresholds better than 20 dB HL at frequencies of 250 to 4000 Hz. One CI user was bilaterally implanted. There was one CI user with some residual hearing (only on 250 Hz). This CI user normally wears a hearing aid, but did not use this hearing aid during testing. For all other CI users, thresholds were over 60 dB on both ears. Therefore, the other CI users did not have access to any residual hearing that would have interfered with our experiment. Duration of deafness for the CI users ranged from 15 until 23 years. Table 2 shows the demographics of the CI participants.

Table 2. CI participant demographics.

CI participant	Sex	Age	Duration of CI use	CI type	Processor	Manufacturer
1	M	55	9 years	CI24R CS	CP810	Cochlear Ltd.
2	M	69	4 years	HiRes 90K Helix	Harmony	Advanced Bionics Corp.
3	M	54	3 years	HiRes 90K Helix	Harmony	Advanced Bionics Corp.
4	M	63	1 year	CI24RE CA	CP810	Cochlear Ltd.
5	F	65	4 years	HiRes 90K Helix	Harmony	Advanced Bionics Corp.
6	F	69	11 years	CI24R K	CP810	Cochlear Ltd.
7	M	69	2 years	CI24RE CA	CP810	Cochlear Ltd.
8	M	72	4 years	CI24RE CA	Freedom	Cochlear Ltd.
9	F	78	1 year	CI24RE CA	CP810	Cochlear Ltd.
10	F	67	2 years	CI512	CP810	Cochlear Ltd.
11	M	72	5 years	HiRes 90K Helix	Harmony	Advanced Bionics Corp.
12	M	65	3 years	HiRes 90K Helix	Harmony	Advanced Bionics Corp.
13	F	71	2 years	CI24RE CA	CP810	Cochlear Ltd.
14	M	64	8 years	CI24R CA	CP810	Cochlear Ltd.
15	F	28	10 years	CI24R CS	CP810	Cochlear Ltd.
16	F	62	2 years	CI24RE H	CP810	Cochlear Ltd.
17	M	76	1 year	CI24RE CA	CP810	Cochlear Ltd.
18	F	57	1 year	CI24RE CA	CP810	Cochlear Ltd.
19	F	72	9 years	CI24R CA	Freedom	Cochlear Ltd.
20	M	72	9 years	HiRes 90K Helix	Harmony	Advanced Bionics Corp.

The present study is a part of a larger project conducted at the University Medical Center Groningen to identify differences in sound, speech, and music perception between NH musicians and non-musicians, and CI listeners. Therefore, participants largely overlap with the participants in the studies by (Fuller et al. (2014c); Fuller et al. (2014a)) – 17 out of 20 of the CI users – and Fuller et al. (2014b), and as a result they were experienced with behavioral studies. Further, data from the control group of NH listeners in this study overlap with the data from the non-musicians in the study by (Fuller et al. 2014a).

The Medical Ethical Committee of the University Medical Center Groningen approved the study. Detailed information about the study was provided to all participants, and written informed consent was obtained before data collection. A financial reimbursement was provided according to the participant reimbursement guidelines of the Otorhinolaryngology Department.

STIMULI

The recordings used in this study were adjusted from Goudbeek and Broersma's emotion database (Goudbeek and Broersma 2010b; Goudbeek and Broersma 2010a). They recorded a nonce word phrase (/nuto hɔm sɛpikɑŋ/) spoken by eight Dutch speakers (four males; four females) for eight different emotions: 'joy', 'pride', 'anger', 'fear', 'tenderness', 'relief', 'sadness', and 'irritation' (four takes per emotion). In this study we used subsets of these stimuli based on pilot tests with NH listeners (Goudbeek and Broersma, 2010a; 2010b). From the original eight emotions, one emotion was selected for each of the four different categories of the Valence-Arousal matrix (Table 1), namely, 'joy', 'anger', 'relief' and 'sadness', which were also the four best recognized emotions on average in the pilot. For pitch analyses, the two best recognized takes for each of the four emotions and for each of the eight speakers were selected. This resulted in a total of 64 tokens (4 emotions × 8 speakers × 2 takes). For the emotion recognition experiment, a further selection was made. The same two takes of the four emotions were used, but only with the four best recognized speakers (two males; two females): speakers 2, 4, 5, and 6 were selected from the original database for this purpose. This resulted in a total of 32 tokens (4 emotions × 4 speakers × 2 utterances).

ACOUSTIC SIMULATION OF CI

Similar to studies by Fuller et al. (2014a) and Fuller et al. (2014b), acoustic CI simulations were implemented using sine-wave vocoded simulations based on a Continuously Interleaved Sampling (CIS) strategy (Wilson, Finley, and Lawson 1991) with Angelsound Software™ (Emily Shannon Fu Foundation). No distortion was added in the vocoder. Stimuli were first bandlimited by bandpass-filtering (200-7000 Hz), and then further bandpass-filtered into 8 frequency analysis bands (4th order Butterworth filters with band cutoff frequencies according to the frequency-place formula of Greenwood (1990). For each channel, a sinusoidal carrier was generated, and the frequency of the sine wave was equal to the center frequency of the analysis filter. The temporal envelope was extracted for each channel through lowpass filtering (4th order Butterworth filter with cutoff frequency = 160 Hz and envelope filter slope = 24 dB/octave) and half-wave rectification 192. The amplitude of the modulated sine wave was adjusted to match the RMS energy of the filtered signal. Finally, each band's modulated carriers were summed and the overall level was adjusted to be equal to that of the original recordings. The motivation for using sine wave instead of noise band excitation was that by doing so, the present study's results would be directly comparable to the results of previous studies that similarly investigated effects of reduced pitch cues in CIs (e.g. (Fu, Chinchilla, and Galvin 2004; Fu et al. 2005).

PROCEDURE

Pitch analysis

Using PRAAT (version 5.3.16; (Boersma and Weenink) and a PRAAT script designed to measure F_0 (Hz) and intensity (dB) every 10 milliseconds (Cook 2002), the recordings' pitch content was analyzed via pitch histograms. Incorrect measurements, for example when PRAAT mistakenly interpreted higher formants as F_0 or when the increased energy around 5 kHz of the fricatives [s] was interpreted by PRAAT as F_0, were manually removed based on visual inspection of the histograms. All pitch measurements were subsequently rounded off towards the frequency (Hz) of the nearest semitone. Next, the frequency of each semitone per recording was automatically counted for each recording (see Figure 1 for an example of a pitch histogram, which depicts how many times each fundamental frequency – depicted as semitones – occurred in the respective recording, and hence does not show any higher harmonics of the individual semitones).

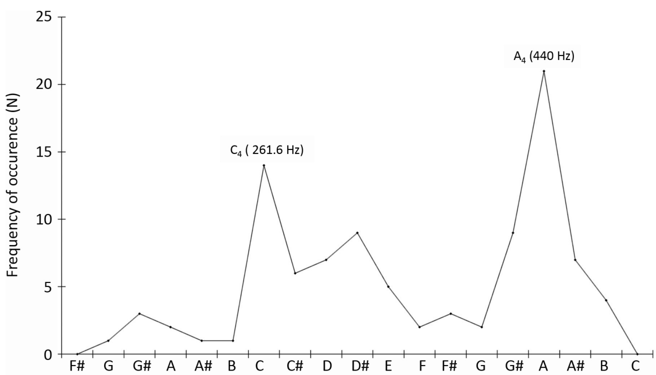


Figure 1. Pitch histogram that shows the semitone frequency for "'Joy', speaker 5, take 3" with dominant pitches indicated.

As this study also investigates the differences in perception of emotions in speech, we assessed mean pitch and pitch range in psychoacoustic scales that characterize how people perceive sound (i.e. in Bark and semitones, respectively) rather than in terms of Hz, the conventional unit of measurement for pitch (Winn, Chatterjee, and Idsardi 2011): since the ratio of frequencies in Hz of two notes exactly an octave apart is always 2:1, the difference between an A_2 note (110 Hz) and an A_3 (220 Hz) is 110 Hz, whereas the difference between an A_3 and an A_4 (440 Hz) is 220 Hz, even though both distances are, according to our auditory perception, exactly the same, namely one octave. For this reason, the pitch measurements in Hz were first converted from Hz into Bark and then averaged per recording prior to data analysis. Our motivation for selecting the Bark scale instead of a log frequency scale (which are in fact rather similar to each other) is that the Bark scale is the scale most suitable for analysis of listeners' perception of formants. Since formants also constitute important acoustic emotion cues we intend to investigate in future studies, selecting the Bark scale for this study already would allow for a better transition into subsequent research (cf. Heeringa, 2004 for a comparison between the Bark scale and frequency scale). The formula used to convert frequencies in Hz (f) into Bark was "Critical band rate (Bark) = ((26.81 × f) / (1960 + f)) – 0.53." If the result of this formula was lower than 2, "0.15 × (2 – result)" was added to the earlier result, and if it was higher than 20.1, "0.22 × (result – 20.1)" was added to the earlier result (Traunmüller 1990). Pitch range for all recordings was first measured in terms of Hz and then converted into semitones prior to data analysis. The number of dominant pitches was assessed from pitch histograms, such as shown in Figure 1. The acoustic signal was considered to have multiple dominant pitches if aside from the most frequent semitone there was another semitone occurring at least half as frequently as the most frequent semitone. Figure 1 shows two distinct dominant pitches for Speaker 5's third take of the high arousal emotion 'joy'.

5

Emotion recognition experiment

All participants were tested in an anechoic chamber. The stimuli were presented using Angelsound Software™ (Emily Shannon Fu Foundation) via a Windows computer (Microsoft) with an Asus Virtuoso Audio Device soundcard (ASUSTeK Computer Inc.). After conversion to an analogue signal via a DA10 digital-to-analog converter (Lavry Engineering Inc.), the stimuli were played via speakers (Tannoy Precision 8D; Tannoy Ltd.) and were presented at 45-80 dB SPL. No masking noise was used. Participants were seated in an anechoic chamber, facing the speaker at a distance of one meter, and they registered their responses to stimuli on an A1 AOD 1908 touch screen (GPEG International).

All participants first completed a training series, and then took part in actual data collection. The NH listeners performed the test two times: first with normal acoustic stimuli and second with CI simulated stimuli. CI users performed the test only once. Each test run lasted around five minutes. The procedure was the same for training and testing, except for two differences. In training, one condition (normal acoustic stimuli) was tested instead of the full set of 32 tokens. Also, in training, feedback was provided. A thumb was shown on the screen in case of a correct answer, or otherwise, both their incorrect and the correct answers were displayed on screen, and were subsequently also played. Participants were presented one randomly selected stimulus at a time, and stimuli were not presented in blocks per speaker. Per auditory stimulus, the participants' task was to indicate on the touchscreen monitor which of the four emotions – 'anger', 'sadness', 'joy', or 'relief' – they heard. Subsequently, percentage scores according to the number of correctly identified tokens were automatically computed.

CI users were instructed to use the normal volume and sensitivity settings of their devices with no further adjustments during the testing.

Ranking of acoustic cues for emotion recognition

To ascertain NH listeners' and CI users' emotion cue rankings, the results of the pitch analyses per speaker were compared to the emotion recognition experiment results per speaker for NH listeners and CI users.

STATISTICAL ANALYSIS

For the statistical analysis, IBM® SPSS® Statistics (version 20) was used. The statistical tests for the pitch analyses were the Mann-Whitney U-test and the independent samples t-test, and for the emotion recognition scores the Kruskal-Wallis test and the Mann-Whitney U-test. A level of $p < .05$ (two-tailed) was considered significant.

RESULTS

Pitch analyses

Mean pitch

Figure 2 shows the mean pitch values (Bark) of the normal acoustic stimuli per emotion with high arousal, low arousal, positive valence, and negative valence indicated. The mean pitch values differed significantly between the four emotions (Kruskal-Wallis test, $\chi^2(3) = 30.399$, $p < .001$). The mean pitch of high arousal emotions ('anger' and 'joy') was significantly higher than that of low arousal emotions ('sadness' and 'relief') (Mann-Whitney U-test, $U(n_1=32, n_2=32) = 111.0$, $p < .001$). There was no significant difference regarding mean pitch between positive ('joy' and 'relief') and negative emotions ('anger' and 'sadness') (Mann-Whitney U-test, $U(n_1=32, n_2=32) = 449.0$, $p = .398$). The mean pitch values differed significantly between the eight speakers (Kruskal-Wallis test, $\chi^2(7) = 22.395$, $p < .01$). Furthermore, reflecting the observation that female voices are generally perceived as being higher than male voices, female pitch values in Bark were slightly higher than male ones.

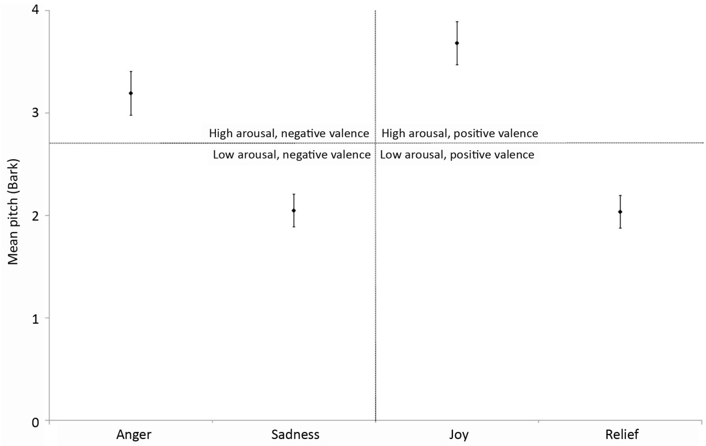


Figure 2. Mean pitch (Bark) – per emotion, averaged from all four speakers and the two takes and with high arousal, low arousal, positive valence, and negative valence indicated; the error bars denote one standard error for this figure and all the following.

Pitch range

Figure 3 shows the average pitch ranges (semitones) of the normal acoustic stimuli per emotion with high arousal, low arousal, positive valence, and negative valence indicated. The findings were similar to those found for the mean pitch: the pitch range values differed significantly between the four emotions (Kruskal-Wallis test, $\chi^2(3) = 28.198$, $p < .001$). Further, high arousal emotions had a significantly wider pitch range in semitones than low arousal emotions (Independent samples t-test, $t(62)=5.944$, $p < .001$). No significant difference regarding pitch range was shown between positive and negative emotions (Mann-Whitney U-test, $t(62)=-.365$, $p = .716$). Furthermore, the pitch range values differed significantly between the eight speakers (Kruskal-Wallis test, $\chi^2(7) = 21.217$, $p < .01$).

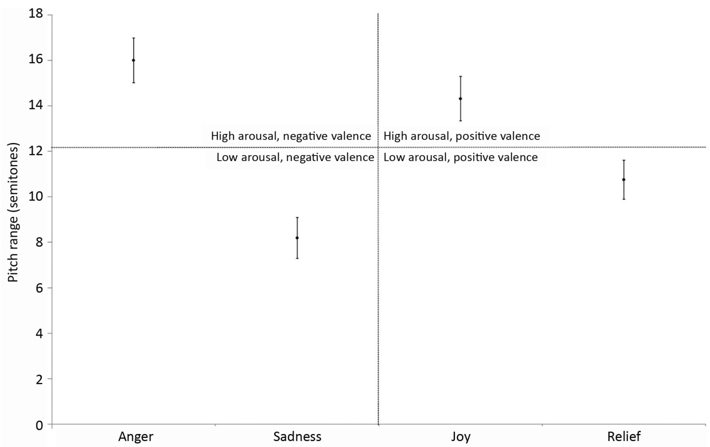


Figure 3. Pitch range (semitones) – per emotion and with high arousal, low arousal, positive valence, and negative valence indicated.

Dominant pitches

Figure 4 shows the number of dominant pitches in the normal acoustic stimuli, averaged per emotion with high arousal, low arousal, positive valence, and negative valence indicated. In contrast to the findings for mean pitch and pitch range, the number of dominant pitches did not differ significantly between the four emotions (Kruskal-Wallis test, $\chi^2(3) = 6.102$, $p = .107$). The number of dominant pitches was significantly higher for high arousal emotions than for low arousal emotions (Mann-Whitney U-test, $U(n_1=32, n_2=32) = 378.5$, $p < .001$). Positive and negative emotions did not differ significantly regarding the number of dominant pitches (Mann-Whitney U-test, $U(n_1=32, n_2=32) = 452.0$, $p = .313$). No significant difference was found among the 8 speakers regarding the number of dominant pitches (Kruskal-Wallis test, $\chi^2(7) = 13.687$, $p = .057$).

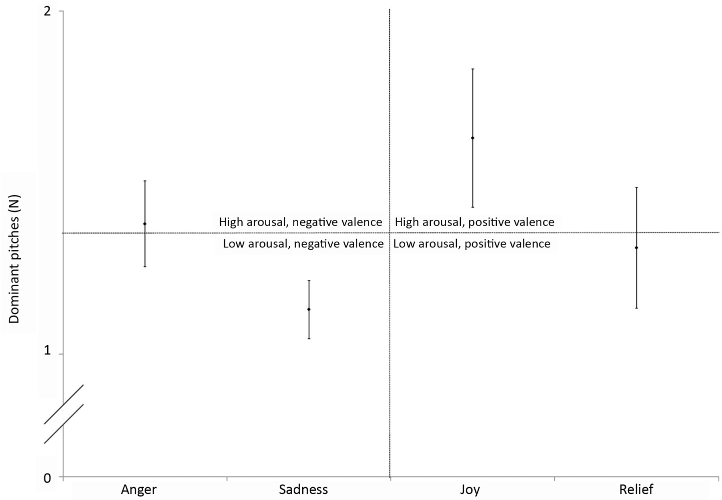


Figure 4. Number of dominant pitches – per emotion with high arousal, low arousal, positive valence, and negative valence indicated.

Emotion recognition experiment

Figure 5 shows the mean emotion recognition for NH participants, for both the normal acoustic stimuli and the CI simulations, as well as for the CI users (with normal acoustic stimuli only). It should be noted here that for the emotion recognition experiment, performance at chance level corresponds to 25% of the emotions being correctly identified. NH participants significantly outperformed CI users with regard to emotion recognition for the normal acoustic stimuli (Mann-Whitney U-test, $U(n_1=20, n_2=20) = .000$, $p < .001$). The NH listeners listening to CI simulations also scored significantly better than the CI users (Independent samples t-test, $t(38) = 6.888$, $p < .001$).

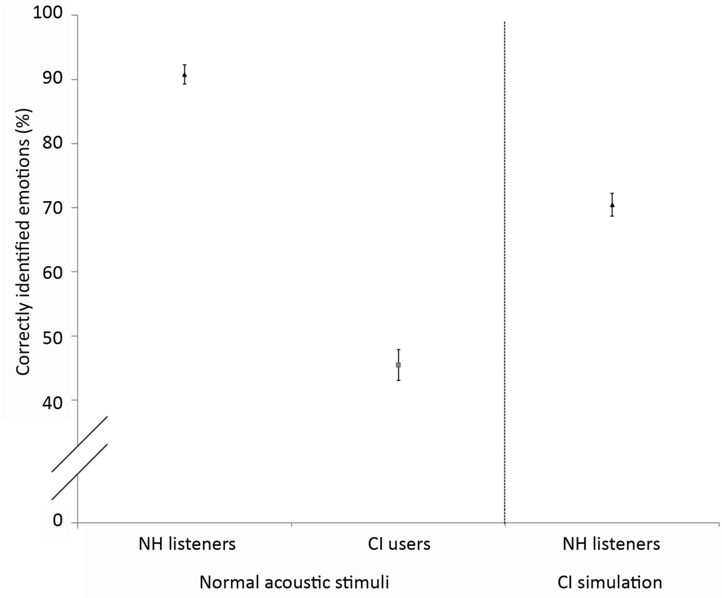


Figure 5. Percentage of correctly identified emotions per condition (normal acoustic stimuli on the left, CI simulations on the right).

NH and CI emotion recognition patterns

Figure 6 shows the recognition of the emotions per speaker for the NH listeners. The recognition of the four speakers' emotions differed significantly for normal acoustic stimuli (Kruskal-Wallis test, $\chi^2(3) = 18.343$, $p < .001$) as well as for CI simulations (Kruskal-Wallis test, $\chi^2(3) = 13.527$, $p < .01$). Post-hoc tests show that NH listeners recognized Speaker 2's emotions significantly worse than all other speakers' emotions for the normal acoustic stimuli and significantly worse than Speaker 4's in the CI simulations (Table 3).

Table 3. Significantly different speaker pairs (Bonferroni corrected) for NH listeners according to a post -hoc analysis using the Mann-Whitney U test.

Condition	Worse speaker	Better speaker	Significance
Emotion	Speaker 2	Speaker 4	$p < .01$
		Speaker 5	$p < .01$
		Speaker 6	$p < .01$
CI simulations	Speaker 2	Speaker 4	$p < .01$

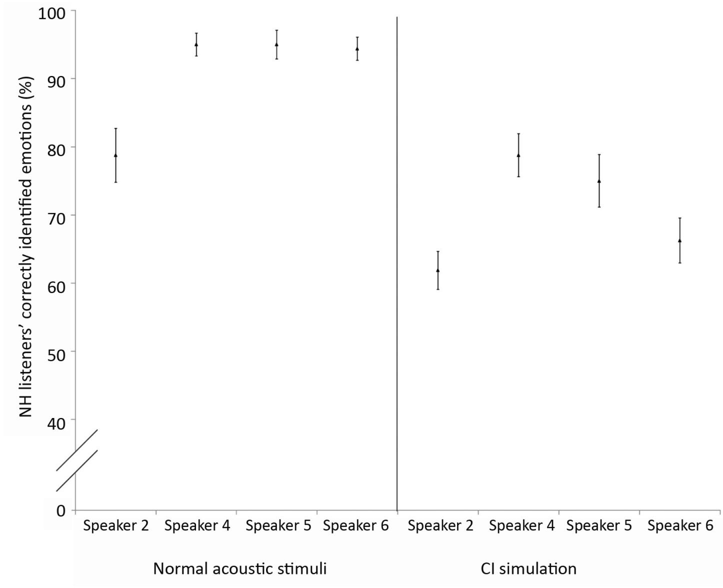


Figure 6. NH listeners' percentage of correctly identified emotions per speaker (normal acoustic stimuli on the left, CI simulations on the right); Speaker 2 and Speaker 4 are male, Speaker 5 and Speaker 6 are female.

Figure 7 shows the recognition of the emotions per speaker for the CI users. Contrary to the NH listeners' recognition scores, the CI users' recognition scores of the four speakers did not differ significantly (Kruskal-Wallis test, $\chi^2(3) = 2.977$, $p = .395$).

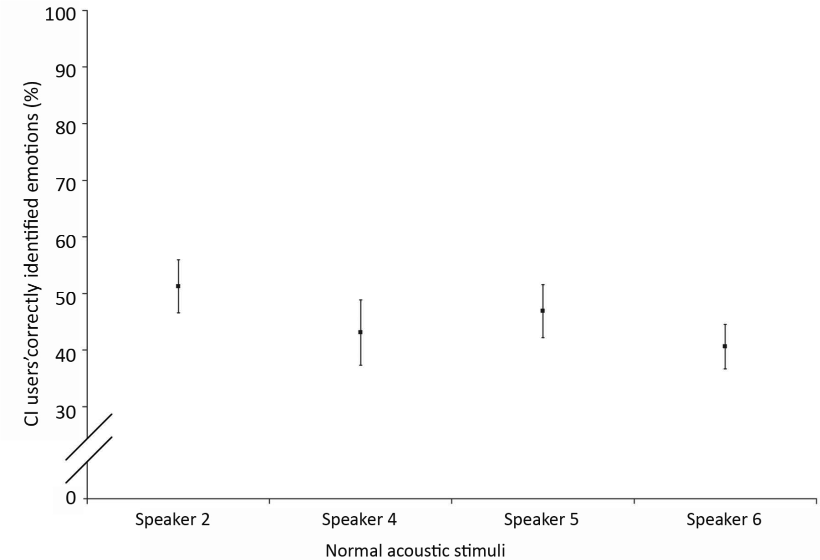


Figure 7. CI users' percentage of correctly identified emotions per speaker (normal acoustic stimuli).

DISCUSSION

Methodology

The present study's aim was to create a framework that accounts for the relative weights of different acoustic pitch cues regarding emotion perception in speech. The method deviates from previous emotion recognition studies in NH listeners and CI users (e.g. Xin, Fu, and Galvin 2007). Firstly, this study is based on a nonce word phrase, which is devoid of any meaning, as opposed to real-language sentences for which it is arguably harder to control whether they are completely semantically neutral. As a result, the possibility that any semantic content of the stimuli would influence participants' emotion recognition could be safely ruled out in the present study. Furthermore, as this study investigates the differences in perception of emotions in speech, we assessed mean pitch and pitch range in psychoacoustic scales that accurately represent how people perceive sound (i.e. in Bark and semitones, respectively) rather than in terms of Hz, the usual unit of measurement for pitch (Winn, Chatterjee, and Idsardi 2011), which was used in previous emotion recognition studies (see e.g. Xin, Fu, and Galvin 2007). In addition, the present study investigated a pitch parameter from the Force of Articulation Model (Gilbers and Van Eerten 2010; Gilbers et al. 2013) that previous CI emotion recognition studies did not, namely the number of dominant pitches occurring in different vocal emotions.

Pitch analyses

The present study's results support its expectations regarding mean pitch (higher mean pitch for high arousal than low arousal emotions), pitch range (wider pitch range for high arousal than low arousal emotions) and number of dominant pitches (more dominant pitches for high arousal than low arousal emotions). The results for mean pitch and pitch range confirm Luo et al.'s (2007) conclusions, and the results for number of dominant pitches confirm the conclusions of Cook (2002), Cook et al. (2004), Schreuder et al. (2006), Liberman (2006), and finally Gilbers and Van Eerten (2010). These studies claim high arousal speech to be characterized by significantly more frequency peaks, i.e. dominant pitches, than low arousal speech. In other words, the number of dominant pitches is a cue for the level of arousal in speech. The results also validate our decision to use the arousal-based force of articulation parameters for the assessment of NH listeners' and CI users' emotion cue rankings. In addition, the results show that regarding pitch-related emotion cues, speakers differentiate between emotions along the Arousal parameter but not along the Valence parameter.

Emotion recognition in NH listeners and CI users

NH listeners were shown to outperform CI users with regard to emotion recognition for the normal acoustic stimuli and also when listening to CI simulations. Please note, however, that because the selected four emotions were the best recognized by NH listeners in the pilot tests, the NH participants might have had an extra advantage compared to the CI users during the emotion recognition experiment. The results can, as a result, not be generalized to all vocal emotions; they are limited to the vocal emotions selected for this study. Moreover, with respect to the role of duration and amplitude cues, it should be noted that when the emotion recognition experiment was conducted, NH listeners and CI users were also tested in conditions with stimuli normalized for duration (1.77 seconds) and amplitude (65 dB SPL). However, since both groups' emotion recognition scores for these conditions were extremely similar to those of the non-normalized conditions (the differences were insignificant across the board), we chose not to include them in our manuscript for reasons of brevity.

Moreover, both groups' recognition scores were compared across speakers to assess whether any speakers' emotions were recognized better or worse than other speakers' – information which could be employed to ascertain NH listeners' and CI users' possibly differing perceptual strategies. In this respect, it was found that Speaker 2's emotions were notably recognized worse in comparison to the other speakers by NH listeners but not by CI users. Even when presented with similar stimuli (CI simulated stimuli for NH listeners and actual CI sound stimuli for CI users), the degree to which emotions were correctly identified differed across speakers for NH listeners (with Speaker 2's emotions being the worst recognized), but it did not for CI users, who recognized each speaker's emotions equally well. In this respect, it should be noted that complete informational equivalence of the stimuli cannot be assumed when comparing emotion recognition in CI simulated stimuli with recognition in actual CI sound stimuli, as simulated CI sound is an approximation of actual CI sound. Nevertheless, the results seem to indicate a difference in strategies for perceiving emotions in speech between NH and CI listeners, as was suggested by Winn et al. (2011). The CI users' differing perceptual strategy is likely due to the degraded cues transmitted with the CI device, but also possibly due to long-term loss of hearing leading to neuroplasticity and exposure to CI-processed sound leading to adaptation. CI users may therefore differ from NH listeners in the relative value they attach to certain emotion cues (e.g. pitch range and mean pitch). In short, due to CI devices' technical limitations and the long term hearing loss with possible loss of neuronal tissue, many acoustic cues are more difficult to perceive for CI users than for NH listeners. Since the NH listeners never had to adjust their perceptual strategies according to CI-like input, they made use of their regular perceptual strategies even in the simulated testing condition, which, as evident from the fact that they recognized one speaker's emotions less often than the others', was not optimal for the CI-simulated input.

Once the difference in perceptual strategies between NH listeners and CI users was established, we analyzed how these differed. Our results indicate that acoustic cues for emotion recognition are ranked by listeners on the basis of salience, and that these cue orderings are different for NH listeners and CI users. This hierarchy is reminiscent of how cues are weighted differently for the recognition of phonemes across languages. Languages differ in the use of perceptual cues for phonetic contrasts. The weights that listeners from different language backgrounds assign to the same cues and for very similar phoneme contrasts can differ strongly, and when listening to a second language, listeners often find it difficult to weigh the cues appropriately, as they tend to pay attention to cues that are important in their own language, and disregard cues that are crucial in the second language (Broersma 2005; Broersma 2010). The acoustic emotion cue hierarchy is also reminiscent of how constraints are ranked differently for different languages within the linguistic framework of Optimality Theory (Prince and Smolensky 2002). Optimality Theory is a non-derivational linguistic theory in which constraints on outputs determine grammaticality. All constraints are universal and available in every language, but languages differ in the way the constraints are ranked in a language-specific dominance hierarchy. In other words, certain constraints are more important in one language than in another. In a similar way, we hypothesize that all acoustic emotion cues are universal and available to all listeners, although due to the limitations in sound transmission in the electrode-nerve interface, cues such as static F_0 levels (mean pitch) are more difficult to discern for CI users than for NH listeners – low F_0 levels are difficult to detect for CI users, but they can still deduce pitch information from the temporal envelope in the signal – and hence CI users have ranked these cues lower in their perceptual strategy than possibly more easily discernible, dynamic F_0 cues (e.g. pitch range). In contrast, since mean pitch is a very robust cue in the acoustic signal, NH listeners, who can easily discern it, place this cue highly on the acoustic emotion cue dominance hierarchy. In the following section, it will be discussed whether we can find evidence for how exactly these cue rankings differ for both groups.

5

Preliminary analysis of cue orderings

In our new approach to acoustic emotion recognition in NH listeners and CI users, which involves the assessment of acoustic emotion cues' relative salience to listeners, two types of information need to be combined in order to ascertain cue rankings: how individual speakers differ from each other in terms of how strongly they distinguish between emotions using pitch, and how they differ from each other concerning how well their emotions are recognized by NH listeners and by CI users. Suppose one particular speaker deviates from the others, for example, because this speaker contrasts between high and low arousal emotions more extremely regarding mean pitch. If this speaker's emotions were recognized better than the other speakers by NH listeners but not by CI users, this would indicate that mean pitch is a relatively important pitch cue for NH listeners and a relatively unimportant one for CI users.

Preliminary analysis of the difference between Speaker 2 (the speaker whose vocal emotions were recognized worse than the other speakers by NH listeners) and the other speakers concerning mean pitch indicates that unlike the other speakers, Speaker 2's mean pitch is below the lowest fundamental frequency one can perceive with a CI device (160 Hz). The fundamental frequency can be perceived with a CI, but only weakly from the spectral and temporal envelope cues. This might suggest that mean pitch is relatively less important for CI users since they did not recognize Speaker 2's emotions any worse than the other speakers.

Regarding pitch range, Speaker 2's high-low arousal emotions ratio is more extreme than the other speakers (roughly 3:1 and 2:1 respectively), and since Speaker 2's emotions were among the better recognized for CI users (note that this difference was not significant) and were the worst recognized by NH listeners (note that this difference was significant), this seems to indicate that CI users attach relatively more value to how much a speaker distinguishes between high and low arousal emotions in terms of pitch range than NH listeners do.

Based on these findings, we hypothesize that for CI users, pitch range is a relatively more salient acoustic emotion cue than mean pitch and is thus ranked higher than mean pitch in their perceptual strategy. In contrast, we hypothesize that for NH listeners, mean pitch is a more salient cue than pitch range, and hence that the former is ranked higher than the latter. Since no speakers deviated from their peers' speech production with respect to the number of dominant pitches, it is thus far not possible to assess the relative rankings of this cue for NH listeners and CI users.

Suggestions for future research

The present study has made use of the speech emotion database recorded by (Goudbeek and Broersma 2010b; Goudbeek and Broersma 2010a), which is based on recordings of the nonce word phrase /nuto hɔm sɛpikɑŋ/. This database already allows for analysis of certain important force of articulation parameters such as speech rate, mean pitch, pitch range, segment duration, syllable isochrony and (lack of) vowel reduction. The Force of Articulation Model (Gilbers et al. 2013), however, consists of other parameters as well, and the nonce word phrase /nuto hɔm sɛpikɑŋ/ does not contain segmental content required for these parameters to be measured. For instance, force of articulation also manifests itself in plosives (in the relatively long duration of their release burst, their occlusion and their voice onset time), liquids (in the relatively high F_2 and low F_1 of /l/), and the extreme vowels /a,i,u/ (in which an expansion of the vowel space can be measured if /a,i,u/ occur in comparable, e.g. stressed, positions). Since not all of these segments are present in the nonce word phrase /nuto hɔm sɛpikɑŋ/, the current database does not allow for analysis of all Force of Articulation Model parameters. Therefore, to assess more extensive emotion cue rankings for CI users and NH listeners than presented in this study, an extended speech emotion database that also allows for the analysis of such additional Force of Articulation Model parameters needs to be recorded.

Furthermore, the present study's results show that regarding pitch-related emotion cues, speakers differentiate between emotions not along the Valence parameter but along the Arousal parameter. However, it remains unclear how they differentiate between positive and negative emotions, and future research needs to be conducted to see which cues play a role in making this distinction.

CONCLUSION

The present study proposes an approach to answer the question whether, and, if so, how CI users and NH listeners differ from each other regarding their perceptual strategies in processing of speech emotions, namely by combining results from phonetic analyses of emotional speech to the results of emotion recognition experiments. The results of the study's pitch analyses and emotion recognition experiment confirm the hypotheses. The fact that CI users' and NH listeners' emotion recognition patterns were significantly different indicates that their perceptual strategies in identifying emotional speech may indeed be different for CI users than for NH listeners. This is likely the result of some acoustic cues being only partially available to CI users and of different degrees of cue sensitivity for CI users and NH listeners. NH listeners' and CI users' relative rankings of the number of dominant pitches cue could not be assessed yet, for all speakers performed similarly regarding this cue. However, it appears that NH listeners and CI users differ from each other regarding which acoustic cues of emotional speech they find more salient. This idea is supported by preliminary analyses suggesting that for NH listeners, mean pitch is a more salient cue than pitch range, whereas CI users are biased towards pitch range cues.

Acknowledgments

We would like to thank the CI users and NH listeners who participated in this study. Furthermore, we would like to thank the research assistants for their help with collecting the data. The fifth author is supported by an otological/neurotological stipendium from the Heinsius-Houbolt Foundation. The sixth author is supported by a Rosalind Franklin Fellowship from the University Medical Center Groningen, University of Groningen, and the VIDI grant 016.096.397 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw). The study is part of the UMCG department's research program 'Healthy Aging and Communication'. There is no conflict of interest regarding this manuscript.

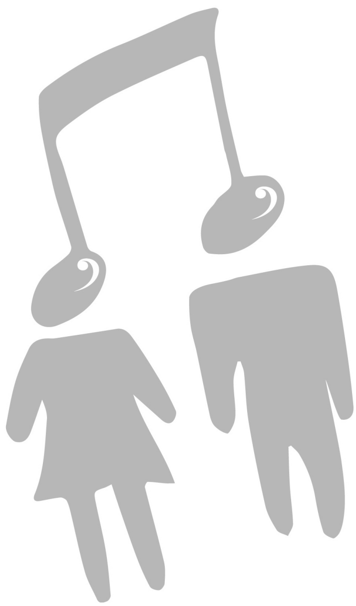
REFERENCES

- Boersma, P., and Weenink, D. PRAAT: Doing phonetics by computer (version 5.3.16). [cited 03-09 2012]. Available from http://www.fon.hum.uva.nl/praat.
- Broersma, M.. 2010. Perception of final fricative voicing: Native and nonnative listeners' use of vowel duration. *The Journal of the Acoustical Society of America* 127 (3): 1636-44.
- Broersma, M.. 2005. Perception of familiar contrasts in unfamiliar positions. *The Journal of the Acoustical Society of America* 117 (6): 3890-901.
- Cook, N. D. 2002. *Tone of voice and mind: The connections between intonation, emotion, cognition and consciousness*. Vol. 47, John Benjamins Publishing.
- Cook, N. D., T. Fujisawa, and K. Takami. 2004. Application of a psychoacoustical model of harmony to speech prosody. Paper presented at Speech Prosody 2004, International Conference, .
- Fitch, H. L., T. Halwes, D. M. Erickson, and A. M. Liberman. 1980. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics* 27 (4): 343-50.
- Francis, A. L., K. Baldwin, and H. C. Nusbaum. 2000. Effects of training on attention to acoustic cues. *Perception & Psychophysics* 62 (8): 1668-80.
- Francis, A. L., N. Kaganovich, and C. Driscoll-Huber. 2008. Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in english. *The Journal of the Acoustical Society of America* 124 (2): 1234-51.
- Friesen, L. M., R. V. Shannon, D. Başkent, and X. Wang. 2001. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America* 110 (2) (Aug): 1150-63.
- Fu, Q. J., S. Chinchilla, and J. J. Galvin III. 2004. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology* 5 (3): 253-60.
- Fu, Q. J., S. Chinchilla, G. Nogaki, and J. J. Galvin III. 2005. Voice gender identification by cochlear implant users: The role of spectral and temporal resolution. *The Journal of the Acoustical Society of America* 118 (3): 1711-8.
- Fu, Q. J., and G. Nogaki. 2005. Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology* 6 (1): 19-27.
- Fu, Q. J., R. V. Shannon, and X. Wang. 1998. Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *The Journal of the Acoustical Society of America* 104 (6): 3586-96.
- Fuller, C. D., J. J. Galvin 3rd, R. H. Free, and D. Başkent. 2014a. Musician effect in cochlear implant simulated gender categorization. *The Journal of the Acoustical Society of America* 135 (3) (Mar): EL159-65.
- Fuller, C. D., E. Gaudrain, J. N. Clarke, J. J. Galvin, Q. J. Fu, R. H. Free, and D. Başkent. 2014b. Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology : JARO* 15 (6) (Dec): 1037-48.
- Gaudrain, E., and D. Başkent. 2015. Factors limiting vocal-tract length discrimination in cochlear implant simulations. *The Journal of the Acoustical Society of America* 137 (3) (Mar): 1298-308.
- Gilbers, D., R. Jonkers, F. Van der Scheer, and J. Feiken. 2013. On the force of articulation in the foreign accent syndrome. In *Phonetics in europe..*, eds. C. Gooskens, R. Van Bezooijen, 11-33. Hamburg: Peter Lang.
- Gilbers, D., and L. Van Eerten. 2010. Musical modality in spontaneous and acted speech. In *Proceedings of the 11th international conference on music perception and cognition (ICMPC 11).*, eds. S. M. Demorest, S. J. Morrison and P. S. Campbell. 118th ed., 360-363Seattle, Washington, USA.
- Goudbeek, Martijn, and Mirjam Broersma. 2010a. The Demo/Kemo corpus: A principled approach to the study of cross-cultural differences in the vocal expression and perception of emotion. Paper presented at 7th International Conference on Language Resources and Evaluation (LREC 2010), .

- Goudbeek, M., and M.Broersma . 2010b. Language specific effects of emotion on phoneme duration. Paper presented at INTERSPEECH.
- House, D.. 1994. Perception and production of mood in speech by cochlear implant users. Paper presented at ICSLP.
- Kovačić, D., and E. Balaban. 2009. Voice gender perception by cochlear implantees. *The Journal of the Acoustical Society of America* 126 (2): 762-75.
- Kovačić, D., and E. Balaban. 2010. Hearing history influences voice gender perceptual performance in cochlear implant users. *Ear and Hearing* 31 (6) (Dec): 806-14.
- Liberman, M. Poem in the key of what. *language log.* 2006. Available from http://itre.cis.upenn.edu/~myl/languagelog/archives/003651.html.
- Massida, Z., P. Belin, C. James, J. Rouger, B. Fraysse, P. Barone, and O. Deguine. 2011. Voice discrimination in cochlear-implanted deaf subjects. *Hearing Research* 275 (1): 120-9.
- Moore, B. C.J., and R. P. Carlyon. 2005. Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In *Pitch.*, 234-277Springer.
- Pereira, C.. 2000. The perception of vocal affect by cochlear implantees. *Cochlear Implants*: 343-5.
- Ponton, C. W., J. J. Eggermont, M. Don, M. D. Waring, B. Kwong, J. Cunningham, and P. Trautwein. 2000. Maturation of the mismatch negativity: Effects of profound deafness and cochlear implant use. *Audiology & Neuro-Otology* 5 (3-4) (May-Aug): 167-85.
- Prince, A. S., and P. Smolensky. 2002. Optimality theory: Constraint interaction in generative grammar.
- Russell, J. A., and A. Mehrabian. 1977. Evidence for a three-factor theory of emotions. *Journal of Research in Personality* 11 (3): 273-94.
- Schreuder, M., L. van Eerten, and D. Gilbers. 2006. Music as a method of identifying emotional speech. Paper presented at The Workshop Programme Corpora for Research on Emotion and Affect Tuesday 23 rd May 2006, .
- Sinnott, J. M., and T. A. Saporita. 2000. Differences in american english, spanish, and monkey perception of thesay-stay trading relation. *Perception & Psychophysics* 62 (6): 1312-9.
- Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America* 88 (1): 97-100.
- Van der Scheer, F., R. Jonkers, and D. Gilbers. 2014. Foreign accent syndrome and force of articulation. *Aphasiology* 28 (4): 471-89.
- Wilkinson, E. P., O. Abdel-Hamid, J. J. Galvin, H. Jiang, and Q. J. Fu. 2013. Voice conversion in cochlear implantation. *The Laryngoscope* 123 (S3): S29-43.
- Wilson, B. S., C. C. Finley, D.T. Lawson et al. 1991. Better speech recognition with cochlear implants. Nature 352:236–238.
- Winn, M. B., M. Chatterjee, and W. J. Idsardi. 2011. The perception of phonetic features and acoustic cues by impaired listeners. *The Journal of the Acoustical Society of America* 130 (4): 2374.
- Xin, Luo, Q. J. Fu, and J. J. Galvin 3rd. 2007. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification* 11 (4) (Dec): 301-15.

Part two:

Musician effect in normal hearing listeners and cochlear-implant users

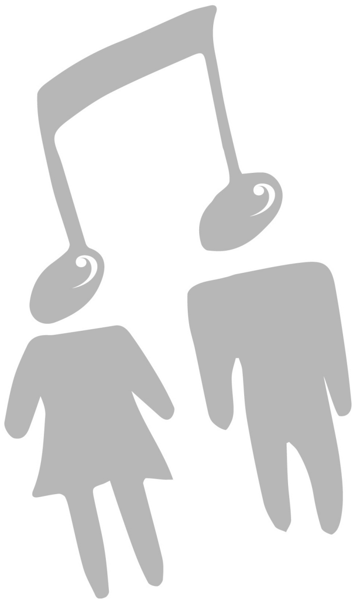


Chapter 6

Musical background not associated with self-perceived hearing performance or speech perception in postlingual cochlear-implant users

Christina Fuller [1,2] Rolien Free [1,2] Bert Maat [1,2] and Deniz Başkent [1,2]

J.Acoust.Soc.Am., 132, 1009-1016.



1 Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
2 Graduate School of Medical Sciences (Research School of Behavioural and Cognitive Neurosciences), University of Groningen, The Netherlands.

ABSTRACT

In normal-hearing listeners, musical background has been observed to change the sound representation in the auditory system and produce enhanced performance in some speech perception tests. Based on these observations, it has been hypothesized that musical background can influence sound and speech perception, and as an extension also the quality of life, by cochlear-implant users.

To test this hypothesis, this study explored musical background [using the Dutch Musical Background Questionnaire (DMBQ)], and self-perceived sound and speech perception and quality of life [using the Nijmegen Cochlear Implant Questionnaire (NCIQ) and the Speech Spatial and Qualities of Hearing Scale (SSQ)] in 98 postlingually deafened adult cochlear-implant recipients. In addition to self-perceived measures, speech perception scores (percentage of phonemes recognized in words presented in quiet) were obtained from patient records.

Self-perceived hearing performance was associated with objective speech perception. Forty-one respondents (44% of 94 respondents) indicated some form of formal musical training. Fifteen respondents (18% of 83 respondents) judged themselves as having musical training, experience, and knowledge. No association was observed between musical background (quantified by DMBQ), and self-perceived hearing-related performance or quality of life (quantified by NCIQ and SSQ), or speech perception in quiet.

INTRODUCTION

Cochlear implants (CIs) are prosthetic devices that restore hearing in profound deafness. Improvements in device design have produced good speech understanding in quiet, but speech perception in noise and enjoyment of music are still not satisfactory (Gfeller et al. 2000b; Leal et al. 2003; Mirza et al. 2003; Kong et al. 2004; McDermott 2004; Kong, Stickney, and Zeng 2005; Lassaletta et al. 2007; Lassaletta et al. 2008; Migirov, Kronenberg, and Henkin 2009; Looi and She 2010). Music is universal, as is language, and is considered the second most important sound processed by humans, after speech (Boucher and Bryden 1997). CI users also rank music, after speech perception, as the second most important acoustical stimulus in their lives (Drennan and Rubinstein 2008). Hence, the improvement of both perception or enjoyment of music can influence the quality of life (QoL) in CI users.

Exposure to music or musical training may also pose specific benefits for sound and speech perception. In normal-hearing (NH) listeners, for example, long-term musical experience can change the sound representation in the auditory system. Enhanced subcortical and cortical representation of speech and brainstem encoding of linguistic pitch are observed with musicians (Musacchia, Strait, and Kraus 2008; Wong et al. 2007; Musacchia et al. 2007). These findings suggest that there may be a shared neural basis for music and language processing (Kraus and Chandrasekaran 2010). Perhaps as a result of this, long-term musically experienced NH adults understand speech in noise better than non-musicians do (Parbery-Clark et al. 2009; Parbery-Clark et al. 2011). Based on these studies with NH musicians, we have hypothesized that musical background might help CI recipients to have better hearing performance and/or speech perception than non-musically trained CI recipients, and thus may increase their health related quality of life (HRQoL).

The effect of musical background on HRQoL and self-perceived hearing-related performance in CI users has been poorly investigated. The study by Lassaletta et al. (2007) is, to the best of our knowledge, the only study that has explored the correlation between musical background and QoL in CI recipients. Two questionnaires, one that evaluates the musical background and the other the QoL, were used in their study. The QoL questionnaire was the Glasgow Benefit Inventory (GBI), a generic questionnaire that measures the patient benefit after otolaryngological interventions and not of the health status per se (Robinson, Gatehouse, and Browning 1996). It is a post intervention questionnaire, not limited to audiological or otolaryngological use only that addresses the benefit after cochlear implantation surgery. In 52 post-lingually deafened CI users, no association between the musical background and the QoL was found. However, a possible reason for the lack of correlation could be the use of a generic patient benefit questionnaire, as this may not have been sufficiently sensitive to reflect hearing functionality-related effects. No analysis was done regarding speech perception measures.

In the present study, we have explored whether musical background has an effect on QoL in postlingually-deafened adult CI users, similar to the study by Lassaletta et al. (2007), but with a number of modifications. We have: 1) employed a larger CI population; 2) used one CI specific HRQoL questionnaire; 3) used one questionnaire specifically developed for hearing-impaired listeners and CI users to assess the self-perceived hearing-related performance, with components on sound and speech perception; 4) additionally analyzed speech perception scores.

MATERIALS AND METHODS

Study population

From the patients implanted and/or monitored at the University Medical Center Groningen (UMCG), 214 CI users were selected and sent the questionnaires based on: current age (older than 18 years), age at the onset of profound hearing loss (5 years or older) and more than one year of CI experience. To include as many patients as possible, etiology and speech perception performance were not used as inclusion criteria. Ninety-eight (46%) replies were received. No significant differences were shown between the respondents and the non-responders for age, CI experience and gender (T-test: $t = -1.038$, $p = 0.301$, $t = -1{,}314$ $p = 0.191$, Chi-square-test: χ^2 0.041, $p = 0.840$, respectively). Among the respondents, one was a bilateral CI recipient. The other demographics of the study participants are shown in the first column of table I.

Dutch Musical Background Questionnaire

The Dutch Musical Background Questionnaire (DMBQ) is a modified and translated version of the Iowa Musical Background Questionnaire (IMBQ; Gfeller et al. (2000b)).[1] Only the first two measurements of the questionnaire assessing the musical background were used. The other parts of the questionnaire are not in the focus of this study as they consider the perception of music with the CI and were therefore excluded from this study.

The first measurement is a musical background score that quantifies formal musical training with questions in six categories: musical instrument lessons, singing lessons, participation in an ensemble, music lessons at elementary school, music lessons at middle school, and music appreciation classes. One point was awarded for each activity that was participated in. Different than the application by Gfeller et al. (2000b), the years of training were left out of the analysis, because many recipients did not know their years of education or were unclear in their answers. Thus, the total scores ranged from 0 (no formal musical training) to 6 (maximum formal musical training), calculated by adding points from all categories. Ninety-four out of 98 respondents completed this section.

The second measurement is a musical background score by self-report in which the respondents rated their musical training, knowledge and experience, in one five-response option question. Hence, each participant had one score varying from 0 (self-report of no

Table I: Demographics of all study participants (first column). The second and the third columns represent the formal musically trained respondents and the respondents without formal musical training based on the first DMBQ measure (N=94).

		Participants	**Formal musical training**	**No formal musical training**
		N = 98	N = 41	N = 53
Gender	Male	39 (40%)	13 (32%)	25 (47%)
	Female	59 (60%)	28 (68%)	28 (53%)
Age (y)		65.6 ± 11.9	61.4 ± 16.6	66.6 ± 11.0
Duration of impaired hearing (y)		37.9 ± 18.6	36.0 ± 19.8	37.0 ± 18.0
CI use since implantation (m)		65.7 ± 33.0	69.6 ± 29.8	64.6 ± 33.9
CI use per day (h)		15.0 ± 2.6	15.6 ± 2.8	15.0 ± 1.7
Education	Lower	13 (14%)	4 (10%)	8 (16%)
	Middle	67 (71%)	26 (65%)	39 (76%)
	Higher	14 (15%)	10 (25%)	4 (8%)
Implant type (no.)	CI22M[a]	1 (1%)	-	1 (2%)
	CI24R CA[a]	24 (24%)	9 (22%)	14 (26%)
	CI24R k[a]	3 (3%)	3 (7%)	-
	CI24RE CA[a]	27 (28%)	13 (32%)	13 (25%)
	CI24R CS[a]	19 (19%)	8 (20%)	11 (21%)
	HiRes90K Helix[b]	24 (24%)	8 (20%)	14 (26%)
Speech processor type (no.)	Esprit3G[a]	31 (32%)	14 (34%)	16 (30%)
	Freedom[a]	43 (44%)	19 (46%)	23 (43%)
	Harmony[b]	24 (24%)	8 (20%)	14 (27%)

a Cochlear Corp., Englewood, Australia device. ACE speech strategy.
b Advanced Bionics Corp., California, USA device. HiRes speech strategy

background) to 4 (self-report of maximum background). Eighty-three out of 98 respondents completed this section.

Nijmegen Cochlear Implant Questionnaire

The Nijmegen Cochlear Implant Questionnaire (NCIQ) is a validated CI-specific HRQoL instrument (Lassaletta et al. 2007; Hinderink, Krabbe, and Van Den Broek 2000). It evaluates not only the self-perceived hearing performance, but also the effects of implantation on the social and psychological functioning and is therefore not a measure of patient benefit, but of HRQoL. The questionnaire has six domains in three categories: physical functioning (sound perception-basic, sound perception-advanced, speech production), social functioning (activity, social functioning) and psychological functioning (self-esteem). The category of physical functioning is a measure of self-perceived hearing performance that evaluates the perception and production of speech and sounds. The second category, social functioning,

6

is a measure of the influence of the hearing impairment on activities and social interactions. The last category, psychological functioning, is a measure of the level of self-esteem of the CI user.

Each domain consists of ten statements, with a five-point response scale indicating the degree to which the statement applies to the respondent. In case of more than three incomplete answers, the corresponding domain is excluded. A total score was calculated by averaging across all 6 domains. The total NCIQ scores ranged from 20 to 88 with a mean of 62 on a 0 to 100 (maximum HRQoL) scale.

Speech, Spatial and Qualities of Hearing Scale

The Speech, Spatial and Qualities of Hearing Scale (SSQ) is a validated environmental and spatial hearing questionnaire, especially developed to range the self-perceived hearing-related abilities in hearing-impaired listeners and CI users (Gatehouse and Noble 2004). It has three subdomains: *speech hearing*, perception of speech in varying scenarios with competing sounds and talkers; s*patial hearing,* judgments of directional hearing and distance; and *other qualities,* assessing segregation of sounds and voices, and listening effort. Note that this questionnaire is a measure of self-perceived hearing-related performance only. Respondents rate themselves with scores varying from 0 to 10 (best). A total score was calculated by averaging scores across all 3 domains. The total SSQ scores ranged from 0 to 7.6 with a mean of 3.5 on a 0 to 10 (maximum hearing performance) scale.

Speech perception in quiet

During regular post-implantation clinic visits, identification of phonemes in recorded consonant-vowel-consonant (CVC) words, presented at 65 and 75 dB SPL (free field) in audiometry booths, is measured with most CI patients (Bosman and Smoorenburg 1995). Percent correct scores are calculated by the ratio of correctly repeated phonemes to the total number of phonemes presented per list at 65 or 75 dB. The highest percent correct score after implantation on either 65 or 75 dB SPL will be used. Scores ranged from 0 to 100%, with a mean of 66%.

Data analysis

Statistical analyses were done using Predictive Analytic Software (PASW) software package version 18.0. Due to non-normality Spearman's correlation coefficient was used to identify and quantify relationships between the scores from DMBQ, NCIQ, SSQ and speech perception. Partial correlation coefficients were conducted between the formal musical training and self-reported musical background and scores of NCIQ, SSQ, and speech perception, corrected for the influence of the educational level, duration of impaired hearing (y), CI use since implantation (m) and CI use per day (h). T-test and Mann-Whitney-U test were conducted for demographic differences between formal musically and non-musically trained and the self-reported musically and non-musically trained. Missing values were excluded by pair and a level of $p<0.05$ (two tailed) was considered significant.

RESULTS

Validation of the SSQ

Table II shows the Spearman's correlation coefficients between the scores per domain and the total score of SSQ and the speech perception score. Significant associations were shown between the domains and total score of the SSQ and the speech perception score.

Table II: Correlations between the speech perception score and the scores of the domains and total score of the SSQ.

	Speech	**Spatial**	**Qualities**	**Total SSQ**
Speech perception score	r = 0.519* (p = 0.000) N = 62	r = 0.483* (p = 0.000) N = 60	r = 0.516* (p = 0.000) N = 60	r = 0.523* (p = 0.000) N = 60

* = Significant

Musical background and HRQoL, self-reported hearing performance and speech perception

Tables III and IV show the results of the first (formal musical training) and second (self-report of musical training, knowledge and experience) measures of DMBQ, respectively. Table V shows the correlation analyses between these measures and the scores of HRQoL (measure of NCIQ) and self-perceived hearing-related performance (measure of SSQ) and speech perception. Figure 1 shows the correlations between the scores of formal musical training (left column) and self-reported musical background (right column) and the total NCIQ score (upper panels), the total SSQ score (middle panels) and speech perception scores (lower panels).

Table III shows the results for formal musical training in 94 respondents. Less than half of the respondents (41 respondents, 44%) had a formal musical training score larger than 0, indicating some form of musical training or lessons taken. The category that was participated in most was musical instrument lessons (26 respondents, 28%; lower part of Table III).

Table IV shows the results for the self-reported musical background in 83 respondents. As seen in the lowest two rows, 15 respondents (18%) judged themselves as having musical training, experience and knowledge.

The main objective of this study was to explore the effect of musical background on speech perception, an objective measure of hearing performance, on HRQoL, as measured by NCIQ, and on self-perceived hearing performance, as measured by SSQ.

Table V and figure 1 show the Spearman's correlation coefficients between the scores of formal musical training and self-reported musical background (left and right columns,

6

Table III. Musical background shown by the first measure of DMBQ.

Formal musical training		Respondents (%)
Formal musical training scores		N = 94 (100%)
Participation in no category	(0 points)	53 (56%)
Participation in one category	(1 point)	11 (12%)
Participation in two categories	(2 points)	19 (20%)
Participation in three categories	(3 points)	8 (9%)
Participation in four categories	(4 points)	1 (1%)
Participation in five categories	(5 points)	2 (2%)
Participation in all six categories	(6 points)	0 (0%)
Participation in each category[a]		N = 94 (100%)
Musical instrument lessons		26 (28%)
Singing lessons		5 (5%)
Participation in musical ensemble		18 (19%)
Music lessons at the elementary school		12 (13%)
Music lessons at the middle school		15 (16%)
Musical theory or appreciation classes		10 (11%)
No formal musical training		53 (56%)

[a] Note that subjects could participate in more than one category

Table IV. Musical background shown by the second measure of DMBQ.

Self-reported musical background	Respondents N = 83 (100%)
No formal training, little knowledge about music, and little experience in listening to music (0 points)	29 (35%)
No formal training or knowledge about music but informal listening experience (1 point)	36 (43%)
Self-taught musician who participates in musical activities (2 points)	3 (4%)
Some musical training, basic knowledge of musical terms, and participation in music classes or ensembles (3 points)	14 (17%)
Several years of musical training, knowledge about music, and involvement in music groups (4 points)	1 (1%)

respectively, in the table, and the left and right panels, respectively, in the figure) and scores of NCIQ, SSQ, and speech perception (top to bottom rows in the table, and top to bottom panels in the figure). The results showed that there were no significant correlations between formal musical training and self-reported musical background and scores of NCIQ, SSQ and speech perception.

Table V. Correlations between the total scores of health-related quality of life (NCIQ), self-perceived hearing performance (SSQ) or speech perception, and the scores of formal musical training and self-reported musical background (first and second measures of DMBQ, respectively).

	Formal musical training		**Self-reported musical background**	
NCIQ	N = 90		N = 79	
	r	p	r	p
Total NCIQ	-0.040	0.708	0.037	0.745
SSQ	N = 75		N = 65	
	r	p	r	p
Total SSQ	-0.194	0.095	-0.030	0.815
Speech perception	N = 70		N = 65	
	r	p	r	p
Speech perception score	0.123	0.311	0.106	0.405

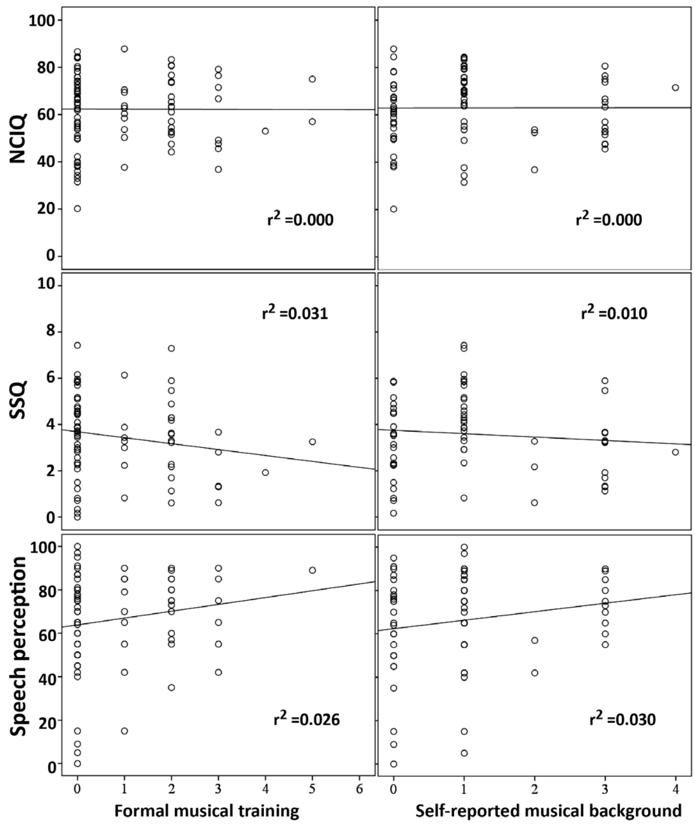


FIGURE 1. Correlations between the scores of formal musical training (left column) and self-reported musical background (right column), and the total scores from NCIQ (upper panels), SSQ (middle panels), and speech perception score (lower panels).

To correct for the influence of the educational level, duration of impaired hearing (y), CI use since implantation (m) and CI use per day (h) on the analyses between NCIQ, SSQ and speech perception and the first two DMBQ measures partial correlation analyses were conducted. Table VI shows the partial correlation coefficients between the scores of formal musical training and self-reported musical background (left and right columns, respectively) and scores of NCIQ, SSQ, and speech perception (top to bottom rows). The results showed that there were no significant correlations.

Table VI. Partial correlations between the scores of health-related quality of life (NCIQ), self-perceived hearing performance (SSQ) or speech perception, and the scores of formal musical training and self-reported musical background (first and second measures of DMBQ, respectively) corrected for the educational level, duration of impaired hearing, CI use since implantation, and CI use per day.

	Formal musical training		**Self-reported musical background**	
NCIQ	N = 81		N = 71	
	r	p	r	p
Total NCIQ	-0.088	0.429	-0.093	0.432
SSQ	N = 63		N = 56	
	r	p	r	p
Total SSQ	-0.217	0.083	-0.117	0.381
Speech perception	N = 62		N= 58	
	r	p	r	p
Speech perception score	0.183	0.147	0.089	0.499

To explore the effect of the musical background in more depth the respondents were divided into musically trained and non-musically trained groups on the basis of the first two measurements of DMBQ shown in table III and IV. The formal musically trained group is the 44% of the 94 respondents that scored 1 or higher on the first DMBQ measure. The self-reported musically trained group is the 18% of the 83 respondents that reported themselves as musically trained. The demographics of the formal musically trained and non-musically trained groups are shown in the second and third columns of table I. The demographics of the self-reported musically trained and non-musically trained are shown in table VII.

Only the distribution of the educational levels was unequal between the groups of formal musically trained and non-musically trained respondents (T-test: $t = -2.005$, $p = 0.049$) and between the self-reported musically trained and non-musically trained (Mann-Whitney-U test: $Z = -3.011$, $p = 0.003$). Tables VIII and IX show the correlations and partial correlations corrected for the educational level between the formal musically trained and self-reported musically trained groups and the total NCIQ and SSQ scores and the speech perception score. No associations between the total NCIQ and SSQ scores or the speech perception score and the formal musically trained and the self-reported musical trained were shown. It must be noted that the number of respondents that reported a musical background was too low to conduct an analyses between the self-reported musically trained and the speech perception score.

Table VII. Demographics of the respondents with a self-reported musical background and without a self-reported musical background based on the second DMBQ measure (N=83).

		Self-reported musical background	**No self-reported musical background**
		N = 15	N = 68
Gender	Male	4 (27%)	30 (44%)
	Female	11 (73%)	38 (56%)
Age (y)		68.5 ± 8.0	62.8 ± 14.9
Duration of impaired hearing (y)		40.4 ± 19.6	35.2 ± 18.9
CI use since implantation (m)		63.3 ± 28.9	67.8 ± 31.8
CI use per day (h)		14.8 ± 3.8	15.4 ± 1.8
Education	Lower	-	9 (13%)
	Middle	8 (57%)	51 (75%)
	Higher	6 (43%)	7 (12%)
Implant type (no.)	CI22M[a]	-	1 (2%)
	CI24R CA[a]	6 (40%)	16 (24%)
	CI24R k[a]	-	2 (3%)
	CI24RE CA[a]	4 (27%)	21 (31%)
	CI24R CS[a]	1 (7%)	14 (20%)
	HiRes 90K 1J[b]	-	-
	HiRes90K Helix[b]	4 (27%)	14 (20%)
Speech processor type (no.)	Esprit3G[a]	5 (33%)	21 (31%)
	Freedom[a]	6 (40%)	33 (49%)
	Harmony[b]	4 (27%)	14 (20%)

a Cochlear Corp., Englewood, Australia device. ACE speech strategy.
b Advanced Bionics Corp., California, USA device. HiRes speech strategy

6

Table VIII. Correlations between the scores of health-related quality of life (NCIQ), self-perceived hearing performance (SSQ) or speech perception, and the respondents with a formal musical training and a self-reported musical background (first and second measures of DMBQ, respectively).

	Formal musical training		**Self-reported musical background**	
NCIQ	N = 41		N = 15	
	r	p	r	p
Total NCIQ	-0.015	0.926	0.247	0.347
SSQ	N = 31		N = 13	
	r	p	r	p
Total SSQ score	-0.301	0.100	-0.077	0.802
Speech perception				
Speech perception score	N = 28			
	r	p		
	0.099	0.614		

Table IX: Partial correlations between the scores of health-related quality of life (NCIQ), self-perceived hearing performance (SSQ) or speech perception, and the respondents with a formal musical training and a self-reported musical background (first and second measures of DMBQ, respectively) corrected for the educational level.

	Formal musical training		**Self-reported musical background**	
NCIQ	N = 37		N = 11	
	r	p	r	p
Total NCIQ	-0.081	0.623	0.287	0.342
SSQ	N = 27		N = 10	
	r	p	r	p
Total SSQ score	-0.264	0.167	0.072	0.823
Speech perception				
Speech perception score	N = 22			
	r	p		
	0.078	0.717		

DISCUSSION AND CONCLUSIONS

In the present study, we have explored musical background in a large population of CI recipients using two measures of DMBQ, one for formal musical training and one for self-reported assessment of musical training. Furthermore, we have explored the correlations and partial correlations of these musical background scores with scores of a CI-specific HRQoL questionnaire, NCIQ, a questionnaire to assess the self-perceived hearing-related performance, SSQ, and phoneme recognition in words in quiet test.

With the first measure of DMBQ, we have observed that 44% of 94 respondents had some formal musical training. Our finding is different than that reported by Leal et al. (2003), who had observed that 62% of 29 participants had some formal musical background. Note that the numbers of the participants (94 vs. 29) and the inclusion criteria in the two studies are vastly different. In our study, we had aimed to have a realistic representation of CI users and therefore had sent the questionnaires to all post-lingually deafened adult CI patients of UMCG, while Leal et al. (2003) had selected patients with very good speech perception performance (all with scores > 75% correct), compared to mean scores in our study of 66%. For the subcategories of musical instrument lessons, participation in a musical ensemble, and musical appreciation classes, we have observed the participation percentages of 28%, 19%, and 11%, respectively. With a similar questionnaire, but again with a smaller number of participants (67), Lassaletta et al. (2008) had observed percentages of 6%, 9% and 22%, respectively, for the same musical subcategories. Philips et al. (2012) found that 20% of 40 CI-recipients followed musical lessons. This variation is not surprising as the results with this measure could vary across different countries and cultures, for example, depending on the mandatory musical training in schools.

With the second measure of DMBQ, the self-reported musical background, we have observed, when we combined the two categories with highest musical training, knowledge, and experience, that 18% of the 83 participants have rated themselves as musically trained. Note that although 44% of the CI recipients had formal musical training, shown by the first DMBQ measure, only 18% have reported themselves as musically trained. One cause for the discrepancy might be the scoring of the first measure of the DMBQ in the present study, where the years of musical training were not taken into account in the formal musical training score. Alternatively, another cause might be that, although a recipient may have played a musical instrument as a child, they may now find themselves, many years later and having been deaf for a long period of time (see Tables I and II), not musically trained. Hence, some CI recipients who had had some previous training might have given up on music now, either due to a long period of deafness or the lack of pleasure in listening to music with the CI, and may not see themselves as musically trained anymore. Our finding of 18% of 83 CI listeners rating themselves as musically trained falls within the results of previous studies; 31% of 65 CI recipients (Gfeller et al. 2000b), 10% of 52 CI recipients (Lassaletta et al. 2007), and 14% of 67 CI recipients (Lassaletta et al. 2008).

Note that in comparison to previous studies our population was, with 94 and 83 participants for the first and second measures, respectively, the largest. Due to this and because we have not pre-selected our patients on performance criteria or etiology of deafness, we argue that our results present a good representation for musical background of typical post-lingually deafened adult CI users.

The main interest of the present study was in the correlations between the DMBQ scores and the HRQoL, self-perceived hearing-related performance or speech perception in quiet. Based on previous studies with NH listeners, assessing the influence of musical training on speech perception and on QoL, (Drennan and Rubinstein 2008; Musacchia, Strait, and Kraus 2008; Wong et al. 2007; Musacchia et al. 2007), we had hypothesized that musical background could be positively correlated with self-perceived hearing-related performance or speech perception performance in CI users, as well as HRQoL. Contrary to our hypothesis, no such association was found. Lassaletta et al. (2007) have similarly found no association between the musical background and the QoL in 52 CI recipients. One potential cause for these comparable findings may be methodological. As both studies showed, the musical background observed in general populations of CI recipients is quite limited, which may make it difficult to produce strong correlations. Focusing only on the CI users that reported a formal musical training or a self-reported musical background matched for all demographic factors and corrected for the differences in educational level, also showed no benefit of musical training on the HRQoL, the self-perceived hearing performance or the speech perception.

Another potential cause for the lack of an association might be the sensitivity of the questionnaires used. Even though we have aimed to use a HRQoL questionnaire specifically prepared for CI users, and a self-perceived hearing-related performance questionnaire specifically prepared for hearing-impaired listeners to assess their own performance of sound and speech perception, it is possible that these questionnaires might still not be sufficiently sensitive. Alternatively, the scoring system used in the DMBQ may not be sufficiently sensitive, as each category participated in was counted as one point, while in reality these categories may have contributed to the musical background in varying levels.

Regarding the (lack of) association between musical background and speech perception in CI recipients, there is no previous study that the present study can be compared to. Our hypothesis on this correlation was based on previous findings with NH populations (Musacchia, Strait, and Kraus 2008; Wong et al. 2007; Musacchia et al. 2007; Kraus and Chandrasekaran 2010; Parbery-Clark et al. 2009; Parbery-Clark et al. 2011). Perhaps the influence of musical training on the auditory system differs between NH listeners and CI users, as damage in the peripheral auditory system can cause the peripheral and central parts of the auditory system to operate differently in hearing impaired listeners compared to normal hearing listeners (Won et al. 2010). Not only the damage to the auditory system on the basis of the etiology of deafness could cause the lack of correlation, also other factors related to CI users, such as duration of deafness, may also have affected the results, but were not taken into account.

Although no associations were found between the musical background and the QoL or speech perception in quiet, we should note that this study was only one measurement in time, and that a relatively small percentage of the study population was musically trained. Hence, this snapshot implies that with no intervention there seems to be no effect of musical background on the QoL or the speech perception within a typical post-lingually deafened adult CI population. However, this finding does not dismiss potential benefits of a systematic musical training program. Focused musical training has been shown to be beneficial in CI users, for example, concerning melodic contour identification (merged), timbre recognition and appraisal (Gfeller et al. 2002), and complex melody task recognition (Gfeller et al. 2000a). This last study by Gfeller et al. (2000a) additionally showed that the CI recipients in the training group also 'liked' music more after training, compared to before. Future research should focus on the effects of such musical training on both adult and pediatric CI population, preferably with longitudinal studies, as the positive effect of musical training in normal-hearing listeners on the perception of speech is influenced by the amount of time that is invested (Parbery-Clark et al. 2009). In addition to potential enhancement of sound or speech perception, such a training program may have other benefits, such as an increase in music appreciation, augmentation of psychosocial wellbeing or development of social skills during group musical therapy. As music is, after all, a significant part of many social and cultural events, the appreciation of music may increase the QoL of CI recipients. In a recent study by Philips et al. (2012) the need of implementing music into the rehabilitation after cochlear implantation was emphasized by CI users themselves, while they believe that musical training might lead to maximal performance with their CI. Therefore, active participation in a musical training program might be of great influence on the QoL of CI recipients.

REFERENCES

- Bosman, A. J., and G. F. Smoorenburg. 1995. Intelligibility of dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology* 34 (5): 260-84.
- Boucher, R., and M. P. Bryden. 1997. Laterality effects in the processing of melody and timbre. *Neuropsychologia* 35 (11) (Nov): 1467-73.
- Drennan, W. R., and J. T. Rubinstein. 2008. Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of Rehabilitation Research and Development* 45 (5): 779-89.
- Galvin, J. J.,3rd, Q. J. Fu, and G. Nogaki. 2007. Melodic contour identification by cochlear implant listeners. *Ear and Hearing* 28 (3) (Jun): 302-19.
- Gatehouse, Stuart, and William Noble. 2004. The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology* 43 (2): 85-99.
- Gfeller, K., S. Witt, J. Stordahl, M. Mehr, and G. Woodworth. 2000a. The effects of training on melody recognition and appraisal by adult cochlear implant recipients. *Journal-Academy of Rehabilitative Audiology* 33 : 115-38.
- Gfeller, K., A. Christ, J. F. Knutson, S. Witt, K. T. Murray, and R. S. Tyler. 2000b. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *Journal of the American Academy of Audiology* 11 (7) (Jul-Aug): 390-406.
- Gfeller, K., S. Witt, M. Adamek, M. Mehr, J. Rogers, J. Stordahl, and S. Ringgenberg. 2002. Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *Journal of the American Academy of Audiology* 13 (3) (Mar): 132-45.
- Hinderink, J. B., P. F. Krabbe, and P. Van Den Broek. 2000. Development and application of a health-related quality-of-life instrument for adults with cochlear implants: The nijmegen cochlear implant questionnaire. *Otolaryngology--Head and Neck Surgery.*123 (6) (Dec): 756-65.
- Kong, Y. Y., G. S. Stickney, and F. G. Zeng. 2005. Speech and melody recognition in binaurally combined acoustic and electric hearing. *The Journal of the Acoustical Society of America* 117 (3): 1351-61.
- Kong, Y. Y., R. Cruz, J. A. Jones, and F. G. Zeng. 2004. Music perception with temporal cues in acoustic and electric hearing. *Ear and Hearing* 25 (2) (Apr): 173-85.
- Kraus, N., and B. Chandrasekaran. 2010. Music training for the development of auditory skills. *Nature Reviews. Neuroscience* 11 (8) (Aug): 599-605.
- Lassaletta, L., A. Castro, M. Bastarrica, R. Perez-Mora, B. Herran, L. Sanz, M. J. de Sarria, and J. Gavilan. 2008. Musical perception and enjoyment in post-lingual patients with cochlear implants. *Acta Otorrinolaringologica Espanola* 59 (5) (May): 228-34.
- Lassaletta, L., A. Castro, M. Bastarrica, R. Perez-Mora, R. Madero, J. De Sarria, and J. Gavilan. 2007. Does music perception have an impact on quality of life following cochlear implantation? *Acta Oto-Laryngologica* 127 (7) (Jul): 682-6.
- Leal, M. C., Y. J. Shin, M. Laborde, M. Calmels, S. Verges, S. Lugardon, S. Andrieu, O. Deguine, and B. Fraysse. 2003. Music perception in adult cochlear implant recipients. *Acta Oto-Laryngologica* 123 (7): 826-35.
- Looi, V., and J. She. 2010. Music perception of cochlear implant users: A questionnaire, and its implications for a music training program. *International Journal of Audiology* 49 (2): 116-28.
- Looi, V., K. Gfeller, and V. Driscoll. 2012. Music appreciation and training for cochlear implant recipients: A review. *Seminars in Hearing* 33 (4) (Nov 1): 307-34.
- McDermott, H. J. 2004. Music perception with cochlear implants: A review. *Trends in Amplification* 8 (2): 49-82.
- Migirov, L., J. Kronenberg, and Y. Henkin. 2009. Self-reported listening habits and enjoyment of music among adult cochlear implant recipients. *Annals of Otology, Rhinology & Laryngology* 118 (5): 350-5.
- Mirza, S., S. A. Douglas, P. Lindsey, T. Hildreth, and M. Hawthorne. 2003. Appreciation of music in adult

patients with cochlear implants: A patient questionnaire. *Cochlear Implants International* 4 (2): 85-95.

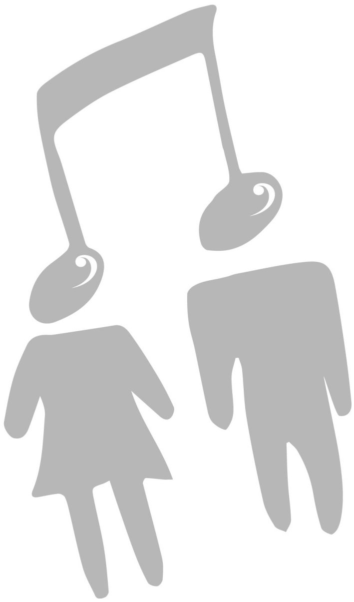
- Musacchia, G., M. Sams, E. Skoe, and N. Kraus. 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America* 104 (40) (Oct 2): 15894-8.
- Musacchia, G., D. Strait, and N. Kraus. 2008. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research* 241 (1-2) (Jul): 34-42.
- Oba, S. I., Q. J. Fu, and J. J. Galvin 3rd. 2011. Digit training in noise can improve cochlear implant users' speech understanding in noise. *Ear and Hearing* 32 (5) (Sep-Oct): 573-81.
- Parbery-Clark, A., D. L. Strait, S. Anderson, E. Hittner, and N. Kraus. 2011. Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. *PLoS One* 6 (5): e18082.
- Parbery-Clark, A., E. Skoe, C. Lam, and N. Kraus. 2009. Musician enhancement for speech-in-noise. *Ear and Hearing* 30 (6) (Dec): 653-61.
- Philips, B., B. Vinck, E. De Vel, L. Maes, W. D'Haenens, H. Keppler, and I. Dhooge. 2012. Characteristics and determinants of music appreciation in adult CI users. *European Archives of Oto-Rhino-Laryngology.*269 (3) (Mar): 813-21.
- Robinson, K., S. Gatehouse, and G. G. Browning. 1996. Measuring patient benefit from otorhinolaryngological surgery and therapy. *Annals of Otology, Rhinology & Laryngology* 105 (6): 415-22.
- Won, J. H., W. R. Drennan, R. S. Kang, and J. T. Rubinstein. 2010. Psychoacoustic abilities associated with music perception in cochlear implant users. *Ear and Hearing* 31 (6) (Dec): 796-805.
- Wong, P. C., E. Skoe, N. M. Russo, T. Dees, and N. Kraus. 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience* 10 (4) (Apr): 420-2.

Chapter 7

Musician effect in cochlear implant simulated gender categorization

Christina D. Fuller [1,2], John J. Galvin III [1,2,3], Rolien H. Free [1,2] and Deniz Başkent [1,2]

J. Acoust. Soc. Am. 135, EL159-65, 2014



1 Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

2 Graduate School of Medical Sciences (Research School of Behavioural and Cognitive Neurosciences), University of Groningen, The Netherlands.

3 Division of Communication and Auditory Neuroscience, House Research Institute, Los Angeles, CA 957; Department of Head and Neck Surgery, David Geffen School of Medicine, UCLA, Los Angeles, CA 995-1624.

ABSTRACT

Musicians have been shown to better perceive pitch and timbre cues in speech and music, compared to non-musicians. It is unclear whether this "musician advantage" persists under conditions of spectro-temporal degradation, as experienced by cochlear-implant (CI) users. In this study, gender categorization was measured in normal-hearing musicians and non-musicians listening to acoustic CI simulations. Dutch words were synthesized to systematically vary fundamental frequency, vocal-tract length, or both to create voices systematically from the female source talker to a synthesized male talker. Results showed an overall musician effect, mainly due to musicians weighting fundamental frequency more in CI simulations.

Introduction

Identifying a talker's gender depends on two anatomically related vocal characteristics: 1) fundamental frequency (F0), mainly related to vocal pitch, and 2) vocal-tract length (VTL), mainly related to the size of the speaker (Smith and Patterson 2005). The ability to identify the voice of a talker is important to separate various talkers in a multi-talker environment and possibly improve speech intelligibility in such situations (Brungart 2001). Recently Fuller et al. (2014) demonstrated that cochlear-implant (CI) users do not utilize both voice cues efficiently. Due to a diminished weighting of VTL cues and an over-reliance on F0 cues, CI users' gender categorization differs from that of normal hearing (NH) listeners, possibly leading to errors in categorization under certain conditions.

NH musicians have been shown to better understand speech in noise, better discriminate voices on the basis of timbre differences, and better perceive pitch in both speech and music, compared to non-musicians (Schon, Magne, and Besson 2004; Chartrand and Belin 2006; Parbery-Clark et al. 2009). This "musician advantage" has been shown to enhance linguistic processing at brainstem, subcortical and cortical levels, and is associated with better functional working memory and auditory attention (e.g. Besson, Chobert, and Marie 2011). Some of the musician advantage for speech-related tasks has been attributed to a better perception of acoustical cues, such as timbre or prosody (e.g. Deguchi et al. 2012). Based on these findings, musicians might be expected to better perceive both F0 and VTL cues compared to non-musicians. Fuller et al. (2014) showed CI users rely almost exclusively on F0 cues to categorize voice gender. NH non-musicians listening to acoustic CI simulations used both cues less efficiently, but relied more strongly on F0 cues in comparison to listening to normal acoustical stimuli. Under such conditions of spectro-temporal degradation, musicians may be better able to extract F0 and VTL information, compared to non-musicians. If so, past musical experience or active music training may benefit CI users' gender categorization that may help speech perception in noise. In this study, voice gender categorization was measured in NH musicians and non-musicians listening to unprocessed speech or to an acoustic CI simulation. Dutch words were synthesized to vary F0, VTL, or both, thereby systematically creating voices from the female source talker to a synthesized male talker. We hypothesized that musicians would be better able to utilize the F0 and VTL cues in a gender categorization task than non-musicians, especially with the CI simulation.

Materials and methods

Participants

Twenty-five NH musicians and twenty-five NH non-musicians were recruited for this study. "Musician" inclusion criteria were defined as: 1) having begun musical training before or at the age of 7 years; 2) having 10 years or more of musical training; and 3) having received some musical training within the last 3 years (Parbery-Clark et al. 2009; Micheyl et al. 2006).

In addition to not meeting the musician criteria, non-musicians were defined as not having received musical training within the 7 years before the study. All participants had pure tone thresholds better than 20 dB HL at audiometric test frequencies between 250 to 4000 Hz, and all were native Dutch speakers with no neurological disorders.

The study was approved by the Medical Ethical Committee of the University Medical Center Groningen. Participants were given detailed information about the study and written informed consent was obtained. A financial reimbursement was provided.

Stimuli

Four meaningful Dutch words in consonant-vowel-consonant format (‘bus’, ‘vaak’, ‘leeg’ and ‘pen’, meaning ‘bus’, ‘often’, ‘empty’, and ‘pencil’, respectively) were used as sources for subsequent speech synthesis. The source speech tokens were taken from the NVA corpus (Bosman and Smoorenburg 1995) and produced by a single, female Dutch talker. The naturally spoken tokens were systematically manipulated to produce voices that ranged from the female to a male talker, using the Straight software (v40.006b), implemented in Matlab and developed by (Kawahara, Masuda-Katsuse, and De Cheveigne 1999). The F0 was decreased by an octave in five steps, 0, 3, 6, 9 or 12 semitones, and the VTL was increased by 23% (resulting in a downward spectral shift of 3.6 semitones) in six steps, 0.0, 0.7, 1.6, 2.4, 3.0 or 3.6 semitones, relative to the female voice. All combinations were generated, resulting in 30 synthesized “voices” and a total of 120 stimuli (4 words x 5 F0 values × 6 VTL values); note that all 120 stimuli were synthesized. The multimedia file contains the word ‘bus’ for: 1) 0 semitone change in F0 and a 0 semitone change in VTL (female voice); 2) 12 semitone change in F0 and 0 semitone change in VTL; 3) 0 semitone change in F0 and 3.6 semitone change in VTL; and 4) 12 semitone change in F0 and 3.6 semitone change in VTL (male voice).

Cochlear implant simulations

Eight-channel, sine-wave vocoded acoustic CI simulations were generated using Angelsound™ software (Emily Shannon Fu Foundation, http://www.angelsound.tigerspeech.com/). The acoustical input was first band-limited to a frequency range of 200-7000 Hz, and then bandpass-filtered into 8 frequency analysis bands [4th order Butterworth filters with band cutoff frequencies according to Greenwood (1990) frequency-place formula]. For each channel, the temporal envelope was extracted using half-wave rectification and lowpass filtering (4th order Butterworth filter with cutoff frequency=160 Hz). These envelopes modulated a sinusoidal carrier that was equal to the center frequency of the analysis filter. The modulated carriers were summed to produce the final stimulus and the overall level was adjusted to be the same level as the original signal. Figure 1 shows the spectra for the word ‘bus’. The middle row shows the original stimulus resynthesized in Straight, with the original parameters of the recorded female voice. In the top row, only the F0 was changed, by an octave down. In the bottom row, only the VTL was changed to be made 23% longer, which results in shifting all the formants down by 3.6 semitones. The left panels show the non-simulated stimulus and the right panel the CI-simulated stimulus.

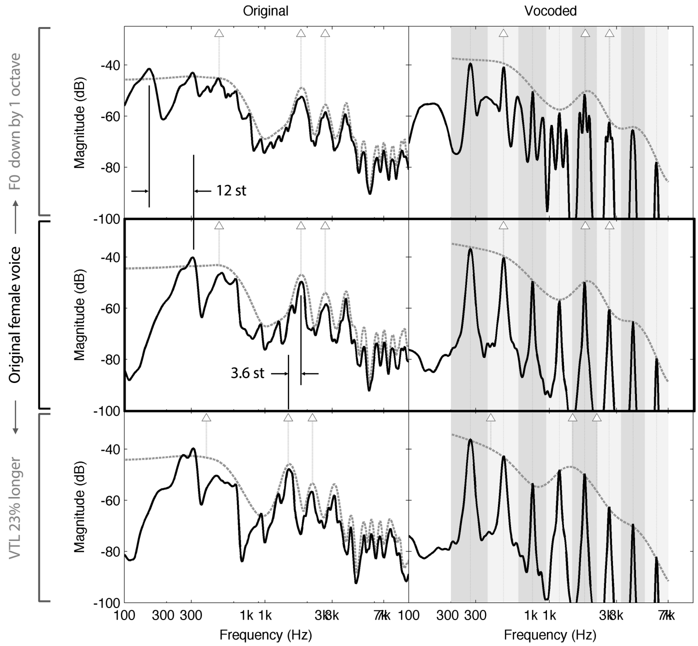


Figure 1. Power spectrum and waveform of the vowel /u/ in 'Bus'. Each row represents a different voice. The *middle row* shows the stimulus resynthesized, in Straight, with the original parameters of the recorded female voice. In the *top row*, only the F0 was changed, by an octave down. In the *bottom row*, only the VTL was changed to be made 23% longer, which results in shifting all the formants down by 3.6 st (semitones). The *left panel* shows the spectra over the duration of the vowel, for the vocoded (*right column*) and non-vocoded (*left column*, noted 'Original') versions of the stimulus. The black solid line represents the spectrum itself, making the harmonics and/or the sinusoidal carriers (and sidebands) of the vocoder visible. The dashed gray line represents the spectral envelope, as extracted by Straight on the left, and interpolating between the carriers for the vocoded sounds on the right. The triangles and stems point to the location of the first three formants, as defined by visual inspection of the Straight envelope, both for the left and the right columns. In the right column, the vocoder analysis filter bands are shown with grayed areas. The frequency of the sine-wave carrier is marked with a dotted line.

Procedure

The stimuli were presented using Angelsound™ software (Emily Shannon Fu Foundation, http://www.angelsound.tigerspeech.com/) and were played from a PC with an Asus Virtuoso Audio Device soundcard (ASUSTeK Computer Inc, Fremont, USA). Participants were seated in an anechoic chamber facing the speaker (Tannoy Precision 8D; Tannoy Ltd., North

Lanarkshire, UK) at one meter distance. After conversion to an analog signal via a DA10 digital-to-analog converter of Lavry Engineering Inc. (Washington, USA), the stimuli were played at 65 dB SPL in the free field. All stimuli were randomly selected from the stimulus set (without replacement) and played to the subject once. The subject indicated whether the talker was a man or woman by selecting one of two response buttons shown on an A1 AOD 1908 touch screen (GPEG International, Woolwich, UK) and labeled “man” or “vrouw” (i.e. “man” and “woman”). Subject responses were recorded by the testing software. No feedback was provided. All of the NH listeners were familiar with CI simulations as they had participated in similar experiments before, but otherwise no specific training for the gender recognition task was provided. The gender categorization task lasted for 10 minutes, resulting in a total testing time of approximately 20 minutes for all participants.

Cue weighting

To quantify how efficiently musicians and non-musicians used voice cues, perceptual weighting of F0 and VTL was calculated using a generalized linear mixed model based on a binomial distribution (logit link function). F0 and VTL were fixed factors and subject was the random intercept. The model was applied to normalized dimensions defined as F0 = -ΔF0/12 and VTL = ΔVTL/3.6, where ΔF0 and ΔVTL represent the F0 or VTL difference in semitones relative to the source talker. With these normalized dimensions, the point (0,0) represents the synthesized female talker and the point (1,1) represents the synthesized male talker. The cue weights were then expressed as *a* and *b* in the equation logit (*score*) = *a* F0 + *b* VTL + ε, where ε is the random intercept that is subject-dependent.

RESULTS

Figure 2 shows the mean voice gender categorization with unprocessed (but synthesized) stimuli for non-musicians (left panels) and musicians (right panels), as a function of F0 difference (top plots) or VTL difference (bottom plots) relative to the female source talker. The percentage of ‘male’ responses was averaged across the four words and across subjects. This is shown by the flatter performance lines and by the lower cue weighting (musicians: F0 2.05, VTL 0.35; non-musicians: F0 1.76, VTL 0.64).

Figure 3 shows similar data sets for musicians and non-musicians as in Figure 2, but with the acoustic CI simulation. In general, both subject groups seemed to use both F0 and VTL less efficiently compared to the unprocessed condition (Figure 2). This is shown by the flatter performance lines and by the lower cue weighting (musicians: F0 2.05, VTL 0.35; non-musicians: F0 1.76, VTL 0.64). The pattern of results with the F0 cue was more diffuse for non-musicians than for musicians, who scored similarly regardless of the VTL cue. The cue weighting analysis suggests that musicians utilized F0 cues more and VTL cues less when compared to non-musicians.

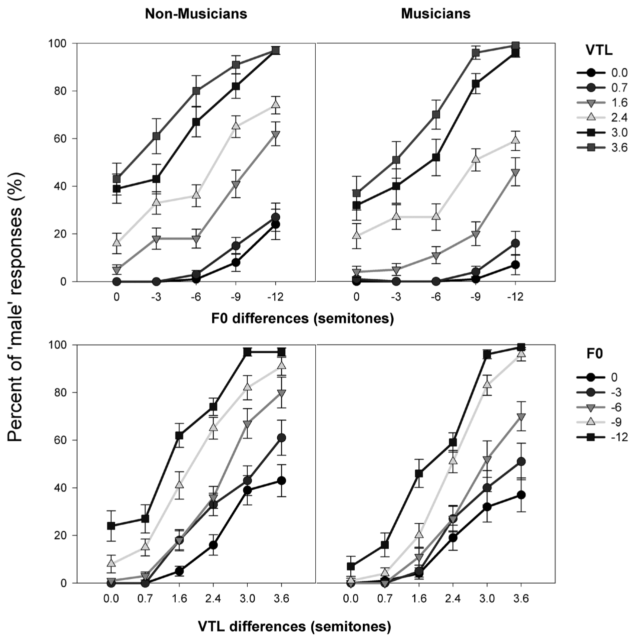


FIGURE 2. Mean gender categorization (across subjects and test words) for non-musicians (left panels) and musicians (right panels) tested with unprocessed (but synthesized) stimuli, as a function of the difference in F0 (top plots) or VTL (bottom plots) relative to the female source talker. Error bars denote one standard error of the mean.

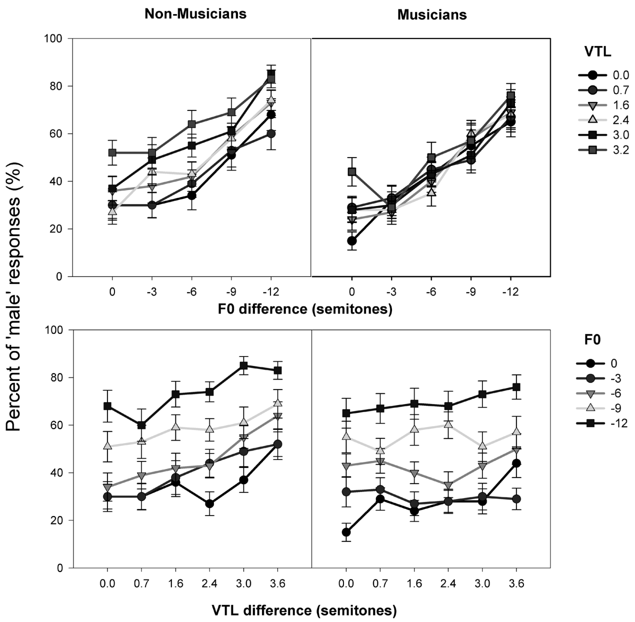


FIGURE 3. Mean gender categorization (across subjects and test words) for non-musicians (left panels) and musicians (right panels) tested with the CI simulation, as a function of the difference in F0 (top plots) or VTL (bottom plots) relative to the female source talker. Error bars denote one error of the mean.

7

A three-way repeated measures, split-plot analysis of variance (RM ANOVA) was performed on all data using a Greenhouser-Geisser correction to correct for sphericity violations (Table 1). The within-subject factors were F0 (5 levels), VTL (6 levels) and listening condition (two levels: unprocessed, CI simulated); the between-subject factor was musical experience (two levels: musician, non-musician). Results confirm a significant overall musician effect on gender categorization. There were significant interactions between F0 and VTL, the listening condition and VTL, and the listening condition, F0 and VTL. However, there were no significant interactions between musical experience and any other factor.

Table 1. Results of split-1 plot, three-way RM ANOVA.

Factors	F-ratio, p-value
Musical experience (musician, non-musician)	F (1,1)= 5.81, p=0.020*
F0	F (2.22,106.73)= 161.94, p<0.001 **
VTL	F (3.27,156.98)= 220.69, p<0.001 **
Listening condition (unprocessed, CI simulation)	F (1,1)= 49.44, p<0.001 **
F0 x Musical experience	F (2.22,106.73)= 0.098, p=0.98
VTL x Musical experience	F (3.27,156.98)= 1.16. p=0.33
Listening condition x Musical experience	F (1,1)= 0.13, p=0.72
F0 x VTL	F (13.33,639.64)= 6.47, p<0.001**
F0 x VTL x Musical experience	F (13.33,639.64)= 1.72, p=0.051
Listening condition x F0	F (1.80,86.32)= 0.82. p=0.43
Listening condition x VTL	F (2.71,129.91)= 125.55, p<0.001**
Listening condition x VTL x Musical experience	F (2.71,129.91)= 1.05, p=0.37
Listening condition x F0 x Musical experience	F (1.80,86.32)= 1.17, p=0.311
Listening condition x F0 x VTL	F (13.14,630.76)= 7.43, p<0.001**
Listening condition x F0 x VTL x Musical experience	F (13.14,630.76)= 1.07. p=0.38

DISCUSSION

Based on previous studies in which a positive musician effect in NH listeners had been observed in speech and music-related tasks (Besson, Chobert, and Marie 2011; Kraus and Chandrasekaran 2010; Patel 2014), we hypothesized that musicians would utilize the voice cues for gender categorization more effectively than non-musicians, especially in spectrally degraded conditions like the CI simulation. This study showed an overall musician effect, mainly in the CI simulation, and that the perceptual weighting of the two voice cues differed between musicians and non-musicians. Musicians perceptually weighted F0 more, but VTL less, than non-musicians in the CI simulation. It is possible that the CI simulation delivered F0 cues more reliably than VTL cues, and musicians made better use of the more reliable cue. Alternatively, musicians may have been more sensitive to F0 cues, and therefore relied on F0 cues more strongly than on VTL cues. If this is the case, musicians would appear to perform similarly to CI users, who have been shown to rely almost exclusively on F0 cues for gender categorization (Fu, Chinchilla, and Galvin 2004). This may be a coincidence, as CI users generally do not have extensive musical experience due to hearing impairment. On the other hand, in CI simulations, as well as in actual CIs, VTL cues are likely less reliable. This is perhaps the reason for the overall low weighting of VTL in CI simulations, by musicians and non-musicians, as well as CI users. Hence, it may be more advantageous to rely on the more robust cue of F0.

While an overall musician effect was observed, note that the perceptual weighting of F0 and VTL cues was similar for musicians and non-musicians with unprocessed speech. Indeed, performance differences were quite small between subject groups with unprocessed speech (Figure 2). Previous studies have shown better voice timbre recognition and pitch perception in both speech and music by musicians listening to unprocessed acoustic stimuli (Chartrand and Belin 2006; Parbery-Clark et al. 2009). The present gender categorization task may have been too easy with unprocessed stimuli, compared to a voice discrimination task (Chartrand and Belin 2006). As such, gender categorization with unprocessed speech may not have been sensitive to musical experience. Furthermore, the F0 steps used to synthesize the present “talkers” may have been too large to elicit differences in performance between musicians and non-musicians observed in previous studies (Chartrand and Belin 2006; Parbery-Clark et al. 2009; Micheyl et al. 2006). Nevertheless, the musician effect on VTL perception or on combined VTL and F0 perception has not been studied previously. The present data do not show a difference between musicians and non-musicians under normal listening situations. The overall findings add to the previous cross-domain effect of musical experience to speech-related tasks, such as voice timbre recognition and gender categorization (Chartrand and Belin 2006). In general, musical experience has been shown to enhance performance in a number of listening tasks. Music training has also been shown to improve CI users’ music and speech perception (Patel 2014; Galvin, Fu, and Nogaki 2007; Looi, Gfeller, and Driscoll 2012; Gfeller et al. 2002; Gfeller et al. 2000). Music training may benefit CI users’ speech perception, especially when pitch cues are important, for example, for separating foreground speech from masking speech and better understanding of speech in multi-talker environments (Brungart 2001).

REFERENCES

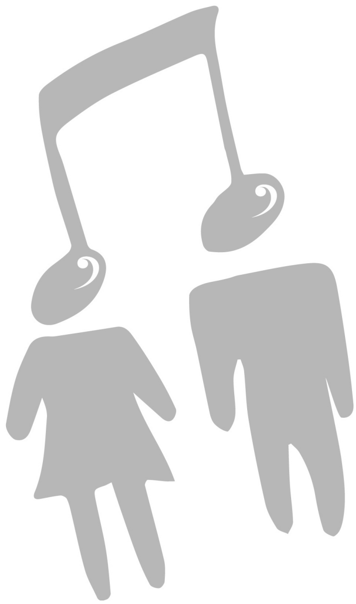
- Besson, M., J. Chobert, and C. Marie. 2011. Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology* 2 (May 12): 94.
- Bosman, A. J., and G. F. Smoorenburg. 1995. Intelligibility of dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology* 34 (5): 260-84.
- Brungart, D. S. 2001. Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America* 109 (3): 1101-9.
- Chartrand, J. P., and P. Belin. 2006. Superior voice timbre processing in musicians. *Neuroscience Letters* 405 (3) (Sep 25): 164-7.
- Deguchi, C., M. Boureux, M. Sarlo, M. Besson, M. Grassi, D. Schön, and L. Colombo. 2012. Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research* 1455 : 75-89.
- Fu, Q. J., S. Chinchilla, and J. J. Galvin III. 2004. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology* 5 (3): 253-60.
- Fuller, C. D., E. Gaudrain, J. N. Clarke, J. J. Galvin, Q. J. Fu, R. H. Free, and D. Başkent. 2014. Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology : JARO* 15 (6) (Dec): 1037-48.
- Galvin, J. J.,3rd, Q. J. Fu, and G. Nogaki. 2007. Melodic contour identification by cochlear implant listeners. *Ear and Hearing* 28 (3) (Jun): 302-19.
- Gfeller, K., S. Witt, J. Stordahl, M. Mehr, and G. Woodworth. 2000. The effects of training on melody recognition and appraisal by adult cochlear implant recipients. *Journal-Academy of Rehabilitative Audiology* 33 : 115-38.
- Gfeller, K., S. Witt, M. Adamek, M. Mehr, J. Rogers, J. Stordahl, and S. Ringgenberg. 2002. Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *Journal of the American Academy of Audiology* 13 (3) (Mar): 132-45.
- Greenwood, D. D. 1990. A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America* 87 (6): 2592-605.
- Kawahara, H., I. Masuda-Katsuse, and A. de Cheveigne. 1999. Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication* 27 (3): 187-207.
- Kovačić, D., and E. Balaban. 2010. Hearing history influences voice gender perceptual performance in cochlear implant users. *Ear and Hearing* 31 (6) (Dec): 806-14.
- Kraus, N., and B. Chandrasekaran. 2010. Music training for the development of auditory skills. *Nature Reviews. Neuroscience* 11 (8) (Aug): 599-605.
- Looi, V., K. Gfeller, and V. Driscoll. 2012. Music appreciation and training for cochlear implant recipients: A review. *Seminars in Hearing* 33 (4) (Nov 1): 307-34.
- Micheyl, C., K. Delhommeau, X. Perrot, and A. J. Oxenham. 2006. Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research* 219 (1-2) (Sep): 36-47.
- Parbery-Clark, A., E. Skoe, C. Lam, and N. Kraus. 2009. Musician enhancement for speech-in-noise. *Ear and Hearing* 30 (6) (Dec): 653-61.
- Patel, A. D. 2014. Can nonlinguistic musical training change the way the brain processes speech? the expanded OPERA hypothesis. *Hearing Research* 308 (Feb): 98-108.
- Schon, D., C. Magne, and M. Besson. 2004. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* 41 (3) (May): 341-9.
- Smith, D. R.R., and R. D. Patterson. 2005. The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and agea). *The Journal of the Acoustical Society of America* 118 (5): 3177-86.

Chapter 8

The musician effect: does it persist in spectrally degraded cochlear implant simulations?

Christina D. Fuller[1,2], John J. Galvin III [1,2,3,4], Bert Maat[1,2],
Rolien H. Free[1,2], Deniz Başkent[1,2]

Frontiers in neuroscience 8 (2014)



1. University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology / Head and Neck Surgery, Groningen, The Netherlands

2. University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Groningen, The Netherlands

3. House Research Institute, Division of Communication and Auditory Neuroscience, Los Angeles, CA, USA

4. Department of Head and Neck Surgery, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA

ABSTRACT

Cochlear implants (CIs) are auditory prostheses that restore hearing via electrical stimulation of the auditory nerve. Compared to normal acoustic hearing, sounds transmitted through the CI are spectro-temporally degraded, causing difficulties in challenging listening tasks such as speech intelligibility in noise and perception of music. In normal hearing (NH), musicians have been shown to better perform than non-musicians in auditory processing and perception, especially for challenging listening tasks. This 'musician effect' was attributed to better processing of pitch cues, as well as better overall auditory cognitive functioning in musicians. Does the musician effect persist when pitch cues are degraded, as it would be in signals transmitted through a CI? To answer this question, NH musicians and non-musicians were tested while listening to unprocessed signals or to signals processed by an acoustic CI simulation. The task increasingly depended on pitch perception: 1) speech intelligibility (words and sentences) in quiet or in noise, 2) vocal emotion identification, and 3) melodic contour identification. For speech perception, there was no musician effect with the unprocessed stimuli, and a small musician effect only for word identification in one noise condition, in the CI simulation. For emotion identification, there was a small musician effect for both. For melodic contour identification, there was a large musician effect for both. Overall, the effect was stronger as the importance of pitch in the listening task increased. This suggests that the musician effect may be more rooted in pitch perception, rather than in a global advantage in cognitive processing (in which musicians would have performed better in all tasks). The results further suggest that musical training before (and possibly after) implantation might offer some advantage in pitch processing that could partially benefit speech perception, and more strongly emotion and music perception.

Keywords: Musician effect, music training, cochlear implant, speech perception, emotion identification, music perception, pitch processing.

INTRODUCTION

In normal hearing (NH), musicians show advantages in auditory processing and perception, especially for challenging listening tasks. Musicians exhibit enhanced decoding of affective human vocal sound (Wong et al. 2007; Musacchia, Strait and Kraus 2008; Strait et al. 2009; Besson, Chobert, and Marie 2011), better perception of voice cues, and better perception of pitch cues in both speech (prosody) and music (Schon, Magne, and Besson 2004; Thompson, Schellenberg, and Husain 2004; Chartrand and Belin 2006). But perhaps more importantly, some transfer of musical training to better speech understanding in noise has also been observed, although evidence for such transfer has been mixed (Parberry-Clark et al. 2009; Kraus and Chandrasekaran 2010; Ruggles et al. 2014). This 'musician effect' might be due to better processing of voice pitch cues that can help to segregate speech from noise (Micheyl et al. 2006; Besson et al. 2007; Oxenham 2008; Deguchi et al. 2012), suggesting that there may be differences between musicians and non-musicians in terms of sound processing at lower levels of the auditory system. Alternatively, the musician effect may be due to better functioning of higher-level processes, such as better use of auditory working memory and attention (Bialystok et al. 2009; Besson, Chobert, and Marie 2011; Moreno et al. 2011; Barrett et al. 2013).

Previously, the musician effect has been studied in NH listeners under conditions in which the spectro-temporal fine structure cues important for complex pitch perception are fully available. It is not yet known if this effect would persist when the acoustic signal is degraded and when the pitch cues are less available, whether due to signal processing and transmission in hearing devices or by hearing impairment. Such is the case with the cochlear implant (CI), the auditory prosthesis for deaf individuals who cannot benefit from traditional hearing aids. Instead of amplifying acoustic sounds, CIs directly stimulate auditory neurons via electrodes placed inside the cochlea. While the CI users can understand speech transmitted through the device to some degree, this speech signal is greatly reduced in spectral resolution and spectro-temporal fine structure. Further, other factors related to electrode-neuron interface may additionally limit CI performance, such as nerve survival patterns (e.g., Başkent and Shannon 2006) or potential mismatch in the frequency-place mapping of electric stimulation (e.g., Başkent and Shannon 2007). As a result, CI users show large variation in their performance for speech perception (Blamey et al. 2013), and most have difficulty understanding speech in noise or in the presence of competing talkers (Friesen et al. 2001; Stickney et al. 2004). The spectro-temporal degradations also severely limit CI users' pitch perception, which is important for recognizing vocal emotion and voice gender, but also for segregating speech from background noise (Fu, Chinchilla, and Galvin 2004; Luo et al. 2007; Oxenham 2008; Fuller et al. 2014b). Problems in pitch processing directly and negatively affect musical pitch and timbre perception, and in turn music perception and appreciation (McDermott 2004; Gfeller et al. 2005; Galvin, Fu, and Nogaki 2007; Heng et al, 2011; Limb and Roy 2014).

Due to aforementioned benefits of the musician effect on speech and music perception, one can argue that music training before or after implantation can also provide some advantages to CI users. In support of this idea, music experience before and after implantation has been shown to benefit CI users' music perception (Gfeller et al. 2000). Further, explicit music training has been observed to significantly improve melodic contour identification (Galvin, Fu, and Nogaki 2007; Galvin et al. 2012), and timbre identification and appraisal (Gfeller et al. 2002; for a review on music appreciation and training in CI users, see Looi, Gfeller, and Driscoll 2012). On a potential connection of music training to speech, however, while some CI studies have shown that better music perception was associated with better speech perception (Gfeller et al. 2007; Won et al. 2010), this connection was not always confirmed by other studies. Fuller et al. (2012) showed that previous musical experience with acoustic hearing did not significantly affect CI users' speech performance after implantation. In that study, as is typical for this patient population, few CI participants were trained musicians before implantation, and many reduced their involvement with music after implantation. It is possible that explicit training after implantation may help postlingually deafened CI users to better associate the degraded pitch patterns via electric hearing to pitch patterns developed during previous acoustic hearing. Alternatively, the spectral degradation with CIs may be so severe that previous music experience provides only limited benefit. Thus, it remains unclear whether the musician effect can persist under conditions of spectro-temporal degradation as experienced by CI users.

Acoustic CI simulations have been widely used to systematically explore signal processing parameters and conditions that may affect real CI users' performance. In a typical CI simulation (e.g., Shannon et al. 1995), the input signal is first divided into a number of frequency analysis bands, then the temporal envelope is extracted from each band and used to modulate a carrier signal (typically band-limited noise or sine-wave), and finally the modulated carrier bands are summed. Parameter manipulations can include the number of spectral channels (to simulate different amounts of spectral resolution), the frequency shift between the analysis and the carrier bands (to simulate different electrode insertions), the envelope filter cut-off frequency (to simulate limits on temporal processing), and the analysis/carrier band filter slopes (to simulate different degrees of channel interaction). CI simulations have also been used to elucidate differences and similarities between acoustic and electric hearing under similar signal processing conditions. Friesen et al. (2001) showed that while NH sentence recognition in noise steadily improved as the number of spectral channels in the acoustic CI simulation increased, real CI performance failed to significantly improve beyond 6-8 channels. Luo et al. (2007) found that temporal envelope cues contributed more strongly to NH listeners' vocal emotion recognition with an acoustic CI simulation than in the real CI case. Kong et al. (2004) showed that NH listeners' familiar melody recognition (without rhythm cues) steadily improved as the number of channels were increased in the CI simulation, while real CI performance remained at chance levels despite having 8-22 channels available in the clinical speech processors.

In the present study, CI simulations were used to differentiate the performance between NH musicians and non-musicians, to identify the effect of long-term musical training, when pitch cues must be extracted from a signal that is spectro-temporally degraded given the limited number of channels. The purpose was two-fold; one, to explore to what degree the musician effect would persist under pitch conditions weakened due to spectro-temporal degradations, and two, to explore if the musician effect could potentially be relevant to CI users perform better with their devices. To achieve this purpose, we systematically investigated the musician effect in a relatively large group of NH participants in three experiments comprised of various speech and music perception tasks, each of which relied on pitch cues to differing degrees. Varying the importance of the pitch cues across the listening tasks might provide insight into mechanisms associated with the musician effect. Speech intelligibility in quiet and in noise was tested using words and sentences. Voice pitch cues would be expected to contribute little to speech understanding in quiet, and possibly more to speech understanding in noise. Vocal emotion identification was tested with and without normalization of amplitude and duration cues that co-vary with fundamental frequency (F0) contours (Luo, Fu, and Galvin 2007; Hubbard and Assmann 2013). Voice pitch cues would be expected to contribute strongly to vocal emotion identification, especially when amplitude and duration cues are less available. Melodic contour identification was tested with and without a competing masker, in which the pitch and the timbre of the masker and target contours were varied. Pitch cues would be expected to contribute most strongly to melodic contour identification, compared to the other listening tasks. All participants were tested in all tasks while listening to unprocessed stimuli or stimuli processed by an 8-channel acoustic CI simulation, using a typical simulation method based on literature. We hypothesized that musicians would exhibit better music perception as a direct result of their musical training. Based on previous studies that showed a transfer from music training to speech perception, we also hypothesized that musicians would better understand speech in noise, and based on previous studies that showed a stronger pitch perception in musicians, to better identify vocal emotion in speech. We further hypothesized that due to better use of pitch cues and better listening skills, musicians would outperform non-musicians also with the CI simulations. However, if musicians outperformed non-musicians in all tasks, this would indicate overall better functioning of high-level auditory perceptual mechanisms. Alternatively, if the musician effect were stronger for listening tasks that relied more strongly on pitch cues, this would indicate that music training mainly improved lower-level auditory perception.

EXPERIMENT 1: SPEECH INTELLIGIBILITY

Rationale

In Experiment 1, we conducted two tests to explore the musician effect on speech intelligibility: 1) word identification in quiet and in noise at various signal-to-noise ratios (SNRs), and 2) sentence identification in various types of noise. In the test of word identification, there was no semantic context, but the words were meaningful; in the test of sentence identification, there was strong semantic context. A musician effect had been previously observed for speech recognition in noise, but with speech materials with intact spectro-temporal fine-structure cues (Parberry-Clark et al. 2009; Kraus and Chandrasekaran 2010). To explore the effect of spectral degradation on speech intelligibility along with the musician effect, NH musicians and non-musicians were tested while listening to unprocessed speech or to an acoustic CI simulation.

MATERIALS AND METHODS

Participants

Twenty-five musicians and twenty-five non-musicians, matched in age and gender, participated in the study (Table 1). Based on previous studies (Micheyl et al. 2006; Parbery-Clark et al. 2009), the inclusion criteria for "musician" were defined as: 1) having begun musical training before or at the age of 7 years, 2) having 10 years or more musical training (i.e., playing an instrument), and 3) having received musical training within the last 3 years on a regular basis. The inclusion criteria for "non-musician" were defined as: 1) not meeting the musician criteria, and 2) not having received musical training within the 7 years before the study. Table 1 shows significant differences between the two participant groups in the number of years of musical training and the starting age of training, confirming a good partition of participants in terms of their music training. There were two small irregularities in participant selection. One non-musician participant started music training at the age of 6 due to mandatory musical training at preliminary school. Another non-musician participant did have 10 years of irregular musical training, but did not have any musical training in the 7 years before the study. Participants were recruited from University of Groningen and from music schools in the area. Further inclusion criteria for all subjects were having normal

Table 1. Demographics of the participants.

	Musicians	Non-musicians	Comparison of the two groups (t-test)
Mean age (range)	22.9 yr (18-27)	22.4 yr (19-28)	t (48) = -0.780; p=0.44
Gender	7 male; 18 female	7 male; 18 female	N.A.
Mean years of musical training (range)	14.6 yr (10-20)	1.6 yr (0-10)	t (48) = -15.96; p<0.001
Mean age of the start of musical training (range)	5.8 yr (3-7)	9.1 yr (6-13)	t (33) = 3.26; p<0.001

hearing (pure tone thresholds better than 20 dB HL at the audiometric test frequencies between 250 to 4000 Hz, and 25 dB HL or better at 8 kHz) and being a native Dutch speaker. Exclusion criteria were neurological disorders, especially dyslexia, psychiatric disorders, or untreated past hearing-related problems.

The Medical Ethical Committee of the University Medical Center Groningen (UMCG) approved the study. Detailed information about the study was given and written informed consent was obtained before participation in the study. A financial reimbursement was provided in line with the guidelines of subject reimbursement of Otorhinolaryngology Department of UMCG.

STIMULI

Word identification

Stimuli included meaningful, monosyllabic Dutch words in CVC format [e.g., bus ('bus,' in English), vaak ('often'), nieuw ('new'), etc.], taken from the NVA test (Bosman and Smoorenburg 1995). The corpus contains digital recordings of twelve lists, each of which contains twelve words spoken by a female talker. Steady speech-shaped noise (provided with the database) that matched the long-term spectrum of the recordings was used for tests conducted with background noise.

Sentence identification

Stimuli included meaningful and syntactically correct Dutch sentences with rich semantic context (Plomp and Mimpen 1979). The corpus contains digital recordings of 10 lists, each of which contains 13 sentences spoken by a female talker. Each sentence contains 4 to 8 words. Sentence identification was measured in three types of noise: 1) Steady speech-shaped noise (provided with the database) that matched the long-term spectrum of the recordings, 2) fluctuating noise, the steady speech-shaped noise additionally modulated by the mean temporal envelope of the sentence recordings, and 3) 6-talker speech babble (Yang and Fu 2005).

Participants were trained with the CI simulation using a different corpus of sentence materials (Versfeld et al. 2000). The training sentences were also meaningful and syntactically correct Dutch sentences with rich semantic context. However, the training sentences were somewhat more difficult compared to the test sentences. The training corpus contains digital recordings of 39 lists, each of which contains 13 sentences spoken by a female talker. Each sentence contains 4 to 9 words.

CI simulation

An acoustic CI simulation was used to replicate some of the spectral and temporal degradations inherent to CI sound transmission (e.g., Shannon et al. 1995). An 8-channel sinewave vocoder based on the Continuous Interleaved Sampling (CIS) strategy (Wilson, Finley, and Lawson 1991) was implemented using Angelsound™ and iStar software (Emily Shannon Fu Foundation, http://www.angelsound.tigerspeech.com/; http://www.tigerspeech.com/istar/istar_about.html). In the simulation, the acoustic input was first band-limited to 200-7000 Hz, which approximates the input frequency range used by many commercial CI devices, and then bandpass-filtered into 8 frequency analysis bands (4th order Butterworth filters with band cutoff frequencies according to Greenwood, 1990, frequency-place formula). Eight channels were used in the CI simulation because previous studies have shown that CI users can only access 6-8 spectral channels (e.g., Friesen et al., 2001). For each channel, the temporal envelope was extracted using half-wave rectification and lowpass filtering (4th order Butterworth filter with cutoff frequency=160 Hz and envelope filter slope = 24 dB/octave). These envelopes were used to modulate a sinusoidal carrier with a frequency that was equal to the center frequency of the analysis filter. The modulated carriers were summed to produce the final stimulus and the overall intensity was adjusted to be the same as the original signal. Figure 1 shows spectrograms for four example Dutch words presented in quiet, for unprocessed speech (left panel) and with the CI simulation (right panel). Similarly, Figure 2 shows spectrograms for an example Dutch sentence presented in quiet, for unprocessed speech (left panel) and with the CI simulation (right panel).

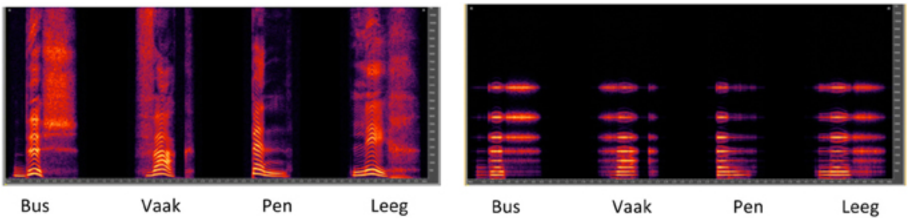


Figure 1. Spectrograms for Dutch monosyllabic words “Bus,” “Vaak,” “Pen,” and “Leeg” (“Bus,” “Often,” “Pen,” and “Empty” in English), shown for unprocessed speech (left panel) or with the CI simulation (right panel).

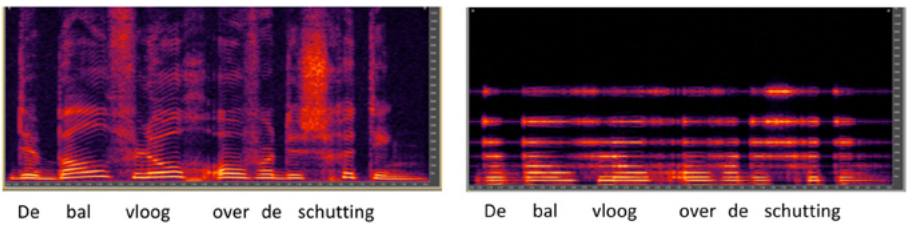


Figure 2. Spectrograms for Dutch sentence “De bal vloog over de schutting” (“the ball flew over the fence” in English), shown for unprocessed speech (left panel) or with the CI simulation (right panel).

Experimental Setup

All tests were conducted in an anechoic chamber. Participants were seated in front of a touchscreen (A1 AOD 1908, GPEG International, Woolwich, UK), facing a loudspeaker (Tannoy precision 8D; Tannoy Ltd., North Lanarkshire, UK) at a distance of 1 meter. Stimuli were presented using iStar custom software (http://tigerspeech.com/istar/) via a Windows computer with an Asus Virtuoso Audio Device soundcard (ASUSTeK Computer Inc, Fremont, USA). After conversion to an analogue signal via a DA10 digital-to-analog converter (Lavry Engineering Inc., Washington, USA) the speech stimulus was played at 65 dBA in free field. The root mean square (RMS) intensity of all stimuli was normalized to the same value. The levels were calibrated with a manikin (KEMAR, GRAS) and a sound-pressure level meter (Type 2610, Brüel Kjær and Sound & Vibration Analyser, Svan 979 from Svantek). Participants' verbal responses on the speech tests were recorded using a DR-100 digital voice recorder (Tascam, California, USA), and were used to double-check responses as needed.

Procedure

The order of the training and testing sessions was the same for all participants. In each experiment, participants received a short training specific to that experiment. The testing was conducted sequentially in this order: word identification, emotion identification, sentence identification, and melodic contour identification. The speech intelligibility data (word and sentence identification) are presented in this section (Experiment 1), the emotion identification data in Experiment 2, and the melodic contour identification data in Experiment 3.

Training

Participants were trained with the CI simulation and in the quiet condition only. Two sentence lists were randomly chosen from the 39 training lists for each participant. The first list was used for passive training, and the second list was used for active training. During passive training, each sentence was played through the loudspeaker and the text was shown simultaneously on the screen. Participants were asked only to listen and to read. After each sentence was presented, the participant pressed '*continue'* on the touchscreen to proceed to the next sentence. After completing the passive training, the touchscreen was turned off. During active training, a training sentence from the second list was played, this time without visual text being displayed. Participants were asked to repeat what they heard as accurately as possible, and to guess if they were unsure of the words. A native Dutch speaker observer, situated in an adjacent room and listening to subjects' responses over headphones, scored the responses using the testing software. Participants were required to score better than 85% correct during active training before beginning formal testing; all participants met this criterion with only one round of active training.

Word identification

Word identification was measured with unprocessed speech and the CI simulation in quiet and in steady, speech-shaped noise at 3 SNRs (+10, +5, and 0 dB). One list of 12 words was used to test each condition (8 lists in total). Word lists were randomly chosen from the 12 lists in the test corpus, and no list was repeated for a participant. The order of the conditions was set to progress from relatively easy to relatively difficult: 1) Unprocessed in quiet, 2) CI simulation in quiet, 3) Unprocessed +10 dB SNR, 4) CI simulation +10 dB SNR, 5) Unprocessed +5 dB SNR, 6) CI simulation +5 dB SNR, 7) Unprocessed 0 dB SNR, and 8) CI simulation 0 dB SNR. During testing, a word was randomly selected from within the list and presented via the loudspeaker. The participant was asked to repeat the word as accurately as possible. The observer listened to the responses and scored each correctly repeated phoneme using testing software that calculated the percentage of phonemes correctly recognized. No trial by trial feedback was provided. The total testing time for all conditions was 12-18 minutes.

Sentence identification

Sentence identification was measured with unprocessed speech and the CI simulation in three types of noise: 1) speech-shaped steady noise, 2) speech-shaped fluctuating noise, and 3) 6-talker babble. One list of 13 sentences was used to test each condition (6 lists in total). Sentence lists were randomly chosen from the 10 lists in the test corpus, and no list was repeated for a participant. Similar to word identification testing, the test order for sentence identification was fixed: 1) Unprocessed in steady noise, 2) CI simulation in steady noise, 3) Unprocessed in fluctuating noise, 4) CI simulation in fluctuating noise, 5) Unprocessed in babble noise, and 6) CI simulation in babble noise. For sentence identification in noise, the speech reception threshold (SRT), defined as the SNR needed to produce 50% correct sentence identification, was measured using an adaptive, one-up/one-down procedure (Plomp and Mimpen 1979), in which the SNR was adjusted from trial to trial according to the accuracy of the response. During testing, speech and noise were presented at the target SNR over the loudspeaker and the participant was asked to repeat the sentence as accurately as possible. If the participant repeated all words in the sentence correctly, the SNR was reduced by 2 dB; if the participant did not repeat all words in the sentence correctly, the SNR was increased by 2 dB. The reversals in SNR between trials 4-13 was averaged and reported as the SRT for the test condition. To better target the SRT within the limited number of sentences in the test list, the initial SNR was different for each noise type and listening condition, based on preliminary testing. For steady noise, the initial SNRs were -4 dB and +2 dB for unprocessed speech and the CI simulation, respectively. For fluctuating noise, the initial SNRs were -8 dB and +6 dB for unprocessed speech and the CI simulation, respectively. For babble, the initial SNRs were -4 dB and +6 dB for unprocessed speech and the CI simulation, respectively. Note that the first sentence was repeated and the SNR increased until the participant repeated the entire sentence correctly. The total testing time for all conditions was 15-20 minutes.

RESULTS

Word identification

Figure 3 shows boxplots for word identification performance by musicians (white boxes) and non-musicians (red boxes) listening to unprocessed stimuli (left panel) or the CI simulation (right panel), as a function of noise condition. Performance generally worsened as the noise level increased, for both listening conditions, and performance with the CI simulation was generally poorer than that with unprocessed stimuli. In the CI simulation, musicians generally performed better than non-musicians. A split-plot repeated measures analysis of variance (RM ANOVA) was performed on the data, with group (musician, non-musician) as the between-subject factor, and listening condition (unprocessed, CI simulation) and SNR (quiet, +10, +5 and 0 dB) as within-subject factors. The complete analysis (with Greenhouser-Geisser corrections due to sphericity violations) is presented in Table 2. There were significant main effects for subject group [**F**(1,48)= 7.76; p =0.008], listening condition [**F**(1,48)= 1098.55; p<0.001] and SNR [**F**(2.63,126.36) = 409.85; p<0.001]. There was a significant interaction between listening condition and SNR [**F**(2.81,134.67)= 148.54; p<0.001]. Despite the overall

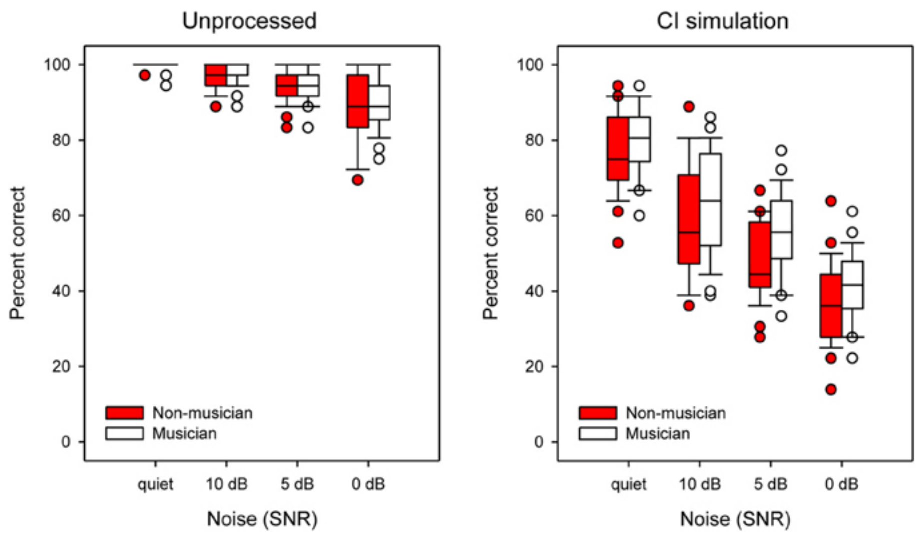


Figure 3. Boxplots of word identification scores for musicians and non-musicians shown as a function of SNR. The left and right panels show data with unprocessed stimuli or with the CI simulation, respectively. The error bars show the 10th and 90th percentiles and the circles show outliers.FIGURE 4. Boxplots of SRTs for musicians and non-musicians shown as function of different noise types for different noise conditions. The left and right panels show data with unprocessed stimuli or with the CI simulation, respectively. The error bars show the 10th and 90th percentiles and the circles show outliers.

main group effect, post-hoc t-tests showed a significant difference between musicians and non-musicians only at the +5 dB SNR with CI simulation [t=-2.94; df=48; p =0.005], namely, at one condition out of eight tested.

Table 2. Results of a split-plot RM ANOVA (with Greenhouse-Geisser correction) for word identification, Experiment 1.

Between-subject factor		**Observed power**
Group	**F**(1,48)= 7.76; p=0.008*	0.78
Within-subject factors		
Listening condition	**F**(1,48)= 1098.55; p<0.001*	1.00
SNR	**F**(2.63,126.36) = 409.85; p<0.001*	1.00
SNR x Listening condition	**F**(2.81,134.67)= 148.54; p<0.001*	1.00
SNR x Group	**F**(2.63,126.36)= 2.02; p=0.50	0.20
Listening condition x Group	**F**(1,48)= 1.02; p=0.051	0.50
Listening condition x SNR x Group	F (2,81, 134,67)=0.95; p=0.42	0.25

* = significant (p<0.05)

Sentence identification

Figure 4 shows boxplots for SRTs by musicians (white boxes) and non-musicians (red boxes) listening to unprocessed stimuli (left panel) or the CI simulation (right panel), as a function of noise type. With unprocessed speech, performance was generally best with the fluctuating noise and poorest with the steady noise. With the CI simulation, performance was generally best with steady noise and poorest with babble. Performance with unprocessed speech was much better than with the CI simulation. Differences between musicians and non-musicians were generally small. A split-plot RM ANOVA was performed on the data, with group as the between-subject factor, and listening condition and noise type (steady, fluctuating, babble) as within-subject factors. The complete analysis is presented in Table 3. There were significant main effects for listening condition [**F**(1,48)= 3771.1; p<0.001] and noise type [**F**(1.56,74.97)= 95.01; p<0.001], but not for group [**F**(1,48)= 2.85; p=0.098]; note that the observed power was relatively weak for the group comparison (0.38). There was a significant interaction between listening condition and noise type [**F**(1.80, 86.54)= 273.90; p<0.001]. Post-hoc tests did not show any significant differences between groups with the different noise types.

EXPERIMENT 2: IDENTIFICATION OF EMOTION IN SPEECH

Rationale

In Experiment 2, a vocal emotion identification task was used to test whether there was a musician effect for a speech-related test that heavily relied on perception of pitch cues in

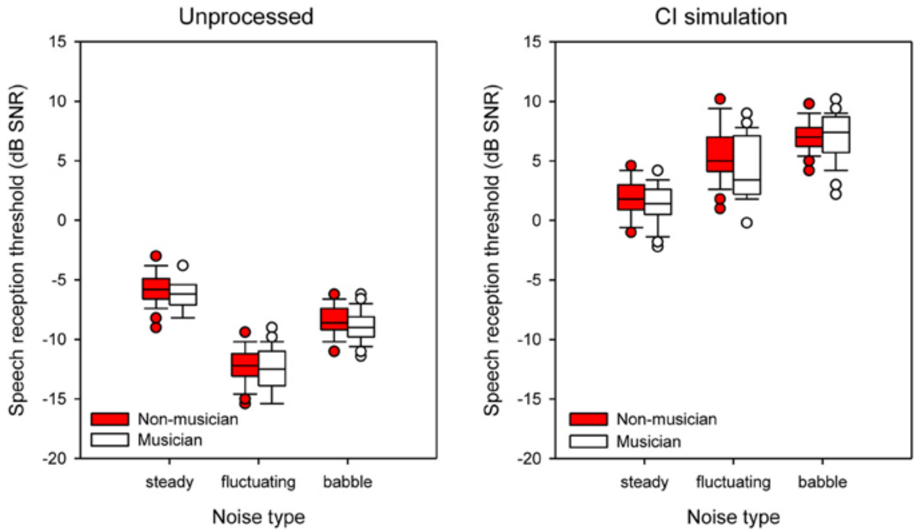


Figure 4. Boxplots of SRTs for musicians and non-musicians shown as function of different noise types for different noise conditions. The left and right panels show data with unprocessed stimuli or with the CI simulation, respectively. The error bars show the 10 and 90th percentiles and the circles show outliers.

Table 3. Results of a split-plot ANOVA (with Greenhouse-Geisser correction) for sentence identification, Experiment 1.

Between-subject factor		**Observed power**
Group	$F(1,48)= 2.85; p=0.098$	0.38
Within-subject factors		
Listening condition	$F(1,48)= 3771.1; p<0.001$*	1.00
Noise type	$F(1.56,74.97)= 95.01; p<0.001$*	1.00
Noise type x Listening condition	$F(1.80, 86.54)= 273.90; p<0.001$*	1.00
Noise type x Group	$F(1.56,74.97)= 0.46; p=0.59$	0.11
Listening condition x Group	$F(1,48)= 0.17; p=0.68$	0.07
Listening condition x Noise type x Group	$F(1.80, 86.54)= 1.05; p=0.35$	0.22

* = significant (p<0.05)

8

speech. To avoid any influence of semantic content on performance, a nonsense word was used to produce the target emotions. Although pitch cues strongly contribute to emotion identification, other cues such as duration and amplitude co-vary with pitch and can also be used for this purpose (Luo, Fu, and Galvin 2007; Hubbard and Assmann 2013). Accordingly, vocal emotion identification was tested for speech stimuli in two versions; once with pitch, duration and amplitude cues preserved across stimuli, and once with duration and amplitude cues normalized across stimuli, leaving in mainly the pitch cues. When duration and amplitude cues are minimal, vocal emotion identification is more difficult, especially under conditions of CI signal processing in which pitch cues are also weakened (Luo et al. 2007). Testing with normalized stimuli would thus allow performance to be compared between musicians and non-musicians when mainly pitch cues are available, with other acoustic cues minimized.

As in Experiment 1, musicians and non-musicians were tested while listening to unprocessed stimuli or to a CI simulation. Participants, CI simulation, and general experimental setup were identical to Experiment 1. The differences in design are explained below.

Stimuli

Stimuli included digital recordings made by Goudbeek and Broersma (2010). The original corpus contains a nonsense word [nutohɔmsɛpikɑŋ] produced by eight professional Dutch actors according to eight target emotions. The actors, who were all trained or were in training at a drama school, were instructed to imagine emotions in a scenario or by reliving personal episodes in which the target emotion occurred. Based on a pilot study with three participants, the four actors (two female, two male) and the four emotions representing all corners of the emotion matrix were chosen for formal testing (Goudbeek and Broersma 2010). Target emotions included: 1) Anger (high arousal, negative valence), 2) Sadness (low arousal, negative valence), 3) Joy (high arousal, positive valence), and 4) Relief (low arousal, positive valence). This resulted in a total of 32 tokens (4 speakers × 4 emotions × 2 utterances).

For the intact stimuli, duration ranged 1.06-2.76 sec and amplitude ranged 45-80 dBA. For the normalized stimuli, duration was normalized to 1.77 sec using a script in PRAAT (version 5.3.16; Boersma and Weenink 2012) without changing the fundamental frequency, and amplitude normalized to 65 dBA using Matlab (i.e., the mean duration and amplitude of the intact stimuli). Figure 5 shows spectrograms for the four target emotions with all cues intact (top panels) or with normalized duration and amplitude cues (bottom panels); the left panels show unprocessed speech and the right panels show speech processed with the CI simulation.

Procedure

For all participants, conditions were tested in a fixed order: 1) Original (with all cues intact), unprocessed stimuli, 2) Original, CI simulation, 3) Normalized (in duration and amplitude), unprocessed, and 4) Normalized, CI simulation. Stimuli were presented using Angelsound software™ (Emily Shannon Fu Foundation, http://www.angelsound.tigerspeech.com/). Before formal testing, participants were familiarized with the test procedure while listening to unprocessed stimuli and to the CI simulation, namely, the target emotions (intact stimuli only) produced by 4 actors not used for formal testing. During training, a target emotion was randomly selected from the stimulus set and presented over the loudspeaker. Subjects were asked to indicate the emotion of the stimulus by touching one of four response boxes on the touchscreen labeled 'anger,' 'sadness,' 'joy,' and 'relief.' Visual feedback was provided on the screen, and in case of an incorrect answer, the correct response and incorrect response were replayed. The actual data collection was identical to training, except that no audio-visual feedback was provided and only the selected test stimuli were used. The software calculated the percent correct and generated confusion matrices. The total testing time for all conditions was 8-16 minutes.

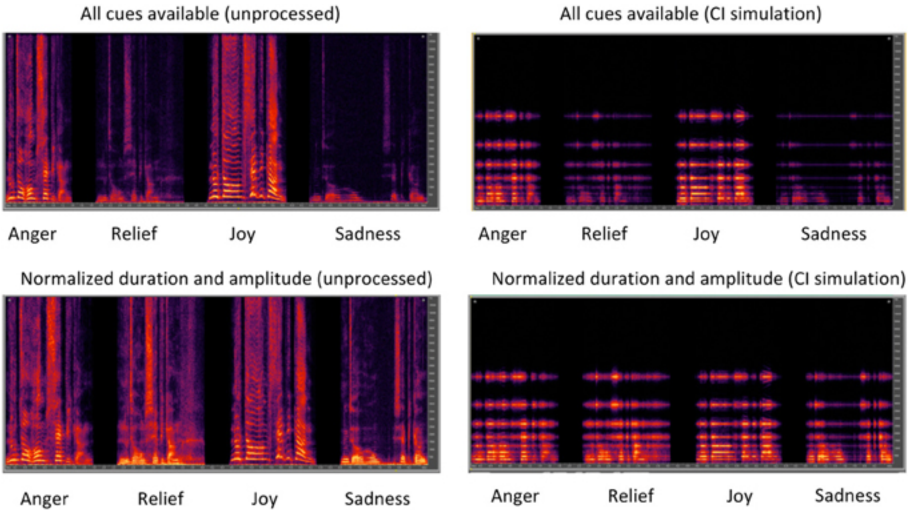


Figure 5. Spectrograms for Dutch nonsense words produced according to four target emotions. The left panels show unprocessed speech and the right panels show speech processed with the CI simulation. The top panels show speech with duration, amplitude, and pitch cues intact. The bottom panels show speech with normalized duration and amplitude cues, but with preserved pitch cues.

RESULTS

Figure 6 shows boxplots for emotion identification by musicians (white boxes) and non-musicians (red boxes) listening to unprocessed stimuli (left panels) or the CI simulation (right panels); the top panels show performance with pitch, duration, and amplitude cues preserved and the bottom panels show performance with normalized duration and amplitude cues. Note that in some cases, median and 25th/75th percentiles could not be displayed because performance was similarly good amongst participants; as a result, only error bars and outliers are displayed. In general, "relief" was the least reliably recognized emotion. Performance generally worsened when duration and amplitude cues were normalized.

There was a small advantage for musicians in all test conditions. A split-plot RM ANOVA

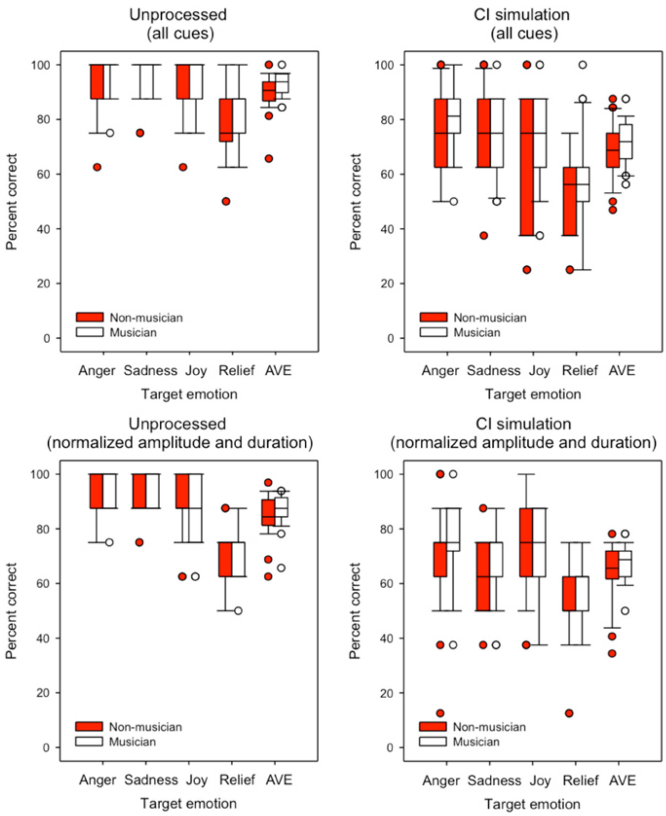


Figure 6. Boxplots for identification of each emotion and overall emotion identification for musicians and non-musicians. The left and right panels show data with unprocessed stimuli or with the CI simulation, respectively. The top panels show performance with pitch, duration, and amplitude cues preserved and the bottom panels show performance with normalized duration and amplitude cues. The error bars show the 10th and 90th percentiles and the circles show outliers.

was performed on the data, with group as the between-subject factor, and listening condition and cue availability (all cues, normalized duration and amplitude) as within-subject factors. The complete analysis is presented in Table 4. There were significant main effects for group [**F**(1,48)= 4.66; p =0.036], listening condition [**F**(1,48)= 323.85; p<0.001] and cue availability [**F**(1,48)= 18.59; p<0.001]. Post-hoc tests showed no significant differences between groups for the different conditions.

EXPERIMENT 3: MELODIC CONTOUR IDENTIFICATION

Rationale

In Experiment 3, a melodic contour identification (MCI; Galvin, Fu, and Nogaki 2007) task was used to test musicians' and non-musicians perception of musical pitch and ability to use timbre and pitch cues to segregate competing melodies. Participants were asked to identify

Table 4. Results of split-plot RM ANOVA for emotion identification, Experiment 2.

Between-subject factor		**Observed Power**
Group	**F**(1,48)= 4.66; p =0. 036*	0.56
Within-subject factors		
Cue availability	**F**(1,48)= 18.59; p<0.001*	0.99
Listening condition	**F**(1,48)= 323.85; p<0.001*	1.00
Cue availability x Group	**F**(1,48)= 0.12; p=0.73	0.06
Listening condition x Group	**F**(1,48)= 0.21; p=0.65	0.07
Cue availability x Listening condition	**F**(1,48)= 1.19; p=0.28	0.19
Cue availability x Listening condition x Group	**F**(1, 48)= 0.030; p=0.86	0.05

* = significant (p<0.05)

a target melodic contour from among a closed-set of responses that represented various changes in pitch direction. MCI was measured for the target alone, and in the presence of a competing contour. The timbre of the target contour and the pitch of the competing contour were varied to allow for different degrees of difficulty in segregating the competing contours. As in Experiments 1 and 2, participants were tested while listening to unprocessed stimuli or the CI simulation. The degradations imposed by CI simulation were expected to have a profound effect on MCI performance, given that melodic pitch was the only cue of interest and would not be well represented in the CI simulation. As this experiment was a more direct measure of music perception, musicians were expected to perform better than non-musicians.

Participants, CI simulation, and general experimental setup were identical to Experiments 1 and 2. Details of the experimental stimuli and procedures are described below.

Stimuli

Stimuli for the MCI test consisted of nine 5-note melodic contours (see Figure 7) that represented different changes in pitch direction: “Rising,” “Flat,” “Falling”, “Flat-Rising,” “Falling-Rising,” “Rising-Flat,” “Falling-Flat,” “Rising-Falling,” “Flat-Falling”). The lowest note in a given contour was A3 (220 Hz). The spacing between successive notes in the contour was 1, 2, or 3 semitones. Presumably, the 1 semitone spacing would be more difficult than the 3 semitone spacing, as the contours would be represented by a smaller cochlear extent. The duration of each note was 250 ms, and the silent interval between notes was 50 ms. The target contour was played by either a piano or an organ sample, as in Galvin, Fu, and Oba (2008). MCI was measured for the target alone or in the presence of a competing contour, as in Galvin, Fu, and Shannon (2009). The competing contour (“masker”) was always the “Flat” contour, played by piano sample. The pitch of the masker was varied to overlap the pitch of

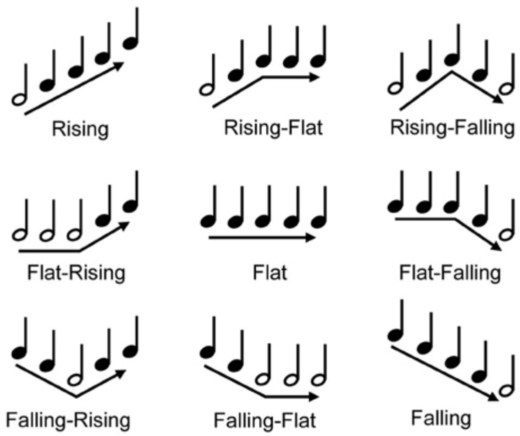


Figure 7. The nine melodic contours used for MCI testing. The white note shows the lowest note of the contour (A3; 220 Hz).

the target, or not. The overlapping pitch was A3 (220 Hz); the non-overlapping pitch was A5 (880 Hz). Thus there were six conditions: 1) piano target alone (no masker), 2) piano target with the A3 piano masker, 3) piano target with the A5 piano masker, 4) organ target alone (no masker) 5) organ target with the A3 piano masker, and 6) organ target with the A5 piano masker. It was expected that MCI performance would be best with no masker, better with the organ than the piano, and better with the A5 than the A3 masker. As such, performance with the organ target with the A5 piano masker (i.e., maximum pitch and timbre difference) would be expected to be better than that with the piano target with the A3 piano masker (minimum pitch and timbre difference). The masker onset and offset was identical to the target contour; thus the notes of the masker and the target occurred simultaneously.

Figure 8 shows spectrograms for the Rising target contour played either by the piano (top panels) or the organ (bottom panels). In each panel, the target contour is shown, from left to right, with no masker, with the overlapping A3 piano masker, and with the non-overlapping A5 piano masker.

Procedure

MCI testing procedures were similar to previous studies (Galvin, Fu, and Nogaki 2007; Galvin, Fu, and Oba 2008; Galvin, Fu, and Shannon 2009). Before formal testing, participants were trained in the MCI procedure. The piano and organ samples were used for training; only the target contours were presented. During training, a contour was randomly selected and presented via the loudspeaker. The participant was instructed to pick the contour that best matched the stimulus from among nine response choices shown on the screen; the response boxes were labeled with both a text descriptor (e.g. “Rising,” Falling”, Flat,” etc.) and an illustration of the contour. After responding, visual feedback was provided and in the

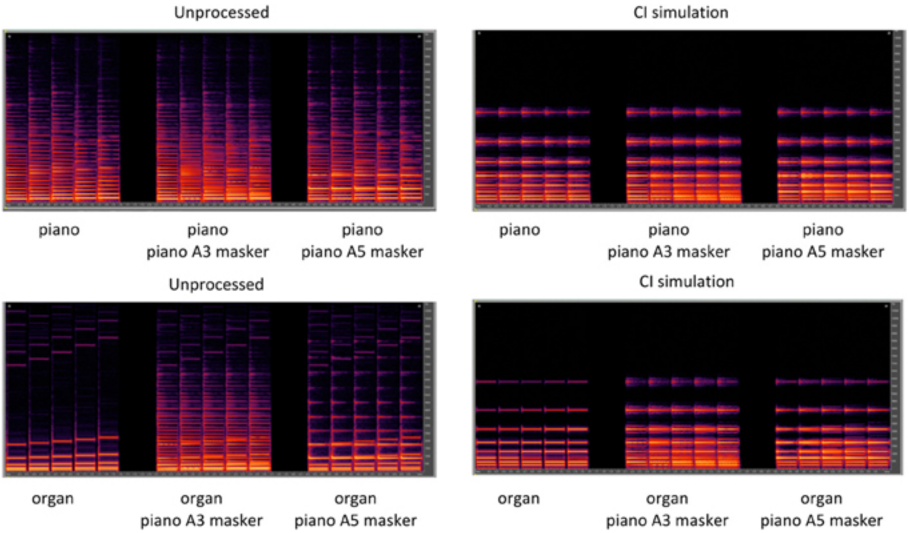


Figure 8. Spectrograms for a Rising target melodic contour with 1-semitone spacing. The top panels show the piano target with the piano A3 and A5 maskers and the bottom panels show the organ target with the piano A3 and A5 maskers. The left panels show unprocessed signals and the right panels show signals processed by the CI simulation.

case of an incorrect response, audio feedback was provided in which the correct response and the participant's (incorrect) response were played in sequence.

Testing methods were the same as for the training, except that no feedback was provided. For all participants, the test order was fixed: 1) piano target (no masker), unprocessed, 2) piano target (no masker), CI simulation, 3) piano target with piano A3 masker, unprocessed 4) piano target with piano A3 masker, CI simulation, 5) piano target with piano A5 masker, unprocessed, 6) piano target with piano A5 masker, CI simulation, 7) organ target (no masker), unprocessed, 8) organ target (no masker), CI simulation, 9) organ target with piano A3 masker, unprocessed, 10) organ target with piano A3 masker, CI simulation, 11) organ target with piano A5 masker, unprocessed and 12) organ target with piano A5 masker, CI simulation. For conditions with a masker, participants were instructed that the masker would always be the "Flat" contour (i.e., the same note played five times in a row), and to ignore the masker and listen for the target, which would change in pitch. Responses were recorded using the test software, and the percent correct was calculated for each condition. The total testing time for all conditions was approximately 30 minutes.

RESULTS

Figure 9 shows box plots of MCI performance with unprocessed stimuli (left panel) or with the CI simulation (right panel), for musicians (white boxes) and non-musicians (red boxes), as a function of test condition. Note that in some cases, median and 25th/75th percentiles

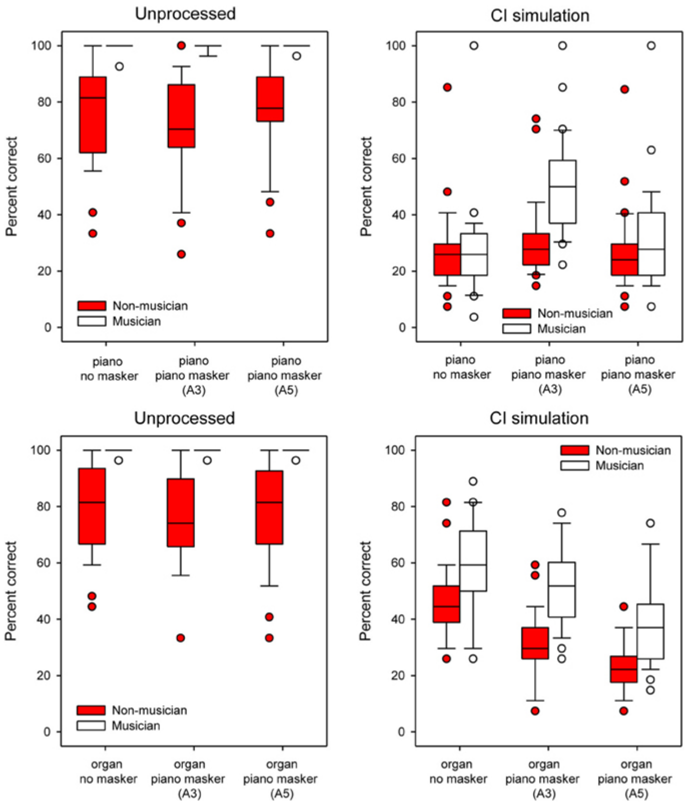


Figure 9. Boxplots for MCI performance for each masker condition for musicians and non-musicians. The top and bottom panels show data for the piano and organ targets, respectively. The left and right panels show data with unprocessed stimuli or with the CI simulation, respectively. Within each panel, data is shown with no masker, with the A3 piano masker, and with the A5 piano masker. The error bars show the 10th and 90th percentiles and the circles show outliers.

could not be displayed because performance was similarly good amongst participants; as a result, only error bars and outliers are displayed. In general, musicians outperformed non-musicians; with unprocessed signals, musician performance was nearly perfect, even with the competing masker. Performance for both groups was much poorer with the CI simulation. The effects of the masker were unclear and somewhat counter-intuitive. In the CI simulation, performance was generally better with the A3 than with the A5 maskers, suggesting that listeners could not make use of the pitch difference between the target and the masker. Similarly, the effects of timbre were small in the CI simulation, as performance was generally similar between the piano and the organ. A split-plot RM ANOVA was performed on the data, with group as the between-subject factor, and target timbre (piano and organ) and masker pitch (no masker A3, A5) as within-subject factors. The complete analysis is presented in Table 5. There were significant main effects for group [**F**(1,48)=59.52; p<0.001], target timbre [**F**(1,48)=69.60; p<0.001], listening condition [**F**(1,48)=993.84; p<0.001], and masker pitch [**F**(1.85,88.71)=14.66; p<0.001]. Post hoc t-tests showed a significant effect of group for all conditions for the unprocessed stimuli (p<0.001). For the CI simulation, a significant musician effect was shown for the piano target with the piano A3 masker [t(48)=-5.10, p<0.001)], the organ target with no masker [t(48)=-2.89, p=0.006], the organ target with the piano A3 masker [t(48)=-5.52, p<0.001] and the organ target with the piano A5 masker [t(48)=-4.22, p<0.001].

Table 5. Results from a split-plot RM ANOVA for melodic contour identification, Experiment 3.

Between subject factor		**Observed power**
Group	**F**(1,48)= 59.52; p<0.001*	1.00
Within subject factors		
Target timbre	**F**(1,48)= 69.60; p<0.001*	1.00
CI simulation	**F**(1,48)= 993.84; p<0.001*	1.00
Masker pitch	**F**(1.85,88.71)= 14.66; p<0.001*	1.00
Target timbre x Masker pitch	**F**(1.76,84.69)= 56.67; p<0.001*	1.00
Target timbre x CI simulation	**F**(1,48)= 55.55; p<0.001*	1.00
Target timbre x Group	**F**(1,48)= 2.90; p=0.10	0.39
CI simulation x Group	**F**(1,48)= 11.19; p=0.002*	0.91
CI simulation x Masker pitch	**F**(1.96,93.82)= 27.51; p<0.001*	1.00
Masker pitch x Group	**F**(1.85,88.71)= 10.45; p<0.001*	0.98
Target timbre x CI simulation x Group	**F**(1,48)= 19.86; p<0.001*	0.99
Target timbre x CI simulation x Masker	**F**(1.95,93.62)=46.22; p<0.001*	0.99
Target timbre x Masker pitch x Group	**F**(1.76,84.69)= 3.21; p=0.051	0.56
Target timbre x CI simulation x Masker pitch x Group	**F**(1.95,93.62)= 1.56; p=0.22	0.32

* = significant (p<0.05)

GENERAL DISCUSSION

The study showed an overall musician effect supporting the hypotheses in general, however, the degree of the musician effect varied greatly across the three experiments. The musician effect was largest for the music test, even with melody contours degraded through a CI simulation, most likely as a direct consequence of music training. The musician effect was smaller for emotion identification, which relied strongly on perception of voice pitch contours, especially for the normalized stimuli where other potential cues, such as intensity and duration, were minimized; however, musicians still outperformed non-musicians even after the pitch cues were also degraded through the CI simulation. For speech perception, there was limited musician effect observed with only one of the speech tests used, word identification, and then only for one out of eight conditions tested, with the CI simulation and presented with background noise at + 5dB SNR.

The musician effect

As outlined in the Introduction, there are two plausible explanations for why musicians may perceive speech better. First, musicians may be better able to detect pitch cues in stimuli, allowing for better segregation of acoustic cues that may improve speech intelligibility in challenging situations (Micheyl et al. 2006; Besson et al. 2007; Oxenham 2008; Deguchi et al. 2012). Second, musicians may be better overall listeners due to better high-level auditory cognitive functioning, such as in working memory and auditory attention (Bialystok et al. 2009; Besson, Chobert, and Marie 2011; Moreno et al. 2011; Barrett et al. 2013), which can also improve speech intelligibility, not only in noise (Parbery-Clark et al. 2009), but also in general. The present data suggest that better pitch processing more strongly contributed to the musician effect, at least for the specific sets of experiments employed. This observation is in line with literature that has shown musicians to rely more heavily on pitch cues than non-musicians when stimuli are degraded (e.g., Fuller et al. 2014a). Further, musicians seem to have a better pitch percept in pitch-related tasks in both speech and music, shown not only behaviorally, but also in imaging studies with an enhanced processing at different brain levels (Besson, Chobert, and Marie 2011). Because it was not explicitly tested in this study, how higher-level cognitive processing may have contributed to the present pattern of results is difficult to judge. However, the observation that the musician effect increased as pitch cues became more meaningful across listening tasks suggests that pitch perception was a strong factor that differentiated musicians from non-musicians.

Prior evidence for transfer of music training to speech perception has been mixed. While Parbery-Clark et al. (2009) showed a small musician effect for identification of sentences presented in noise, but not processed otherwise, Ruggles et al. (2014) showed no musician effect for identification of sentences in noise, presented with or without voice pitch cues. In the present study, there was a significant musician effect for word identification (Exp. 1), yet, this was limited to one condition out of eight tested, only observed in noise and with CI simulation, and there was no musician effect for sentence recognition in noise, with or without CI simulation. The reason for not observing an effect in the latter may be that sentence recognition depends on also other factors besides pitch perception (e.g., segregating speech from noise, extracting meaning with help from semantics, context, prosody, and also using higher-level cognitive and linguistic processes). If the musician effect is largely based on pitch processing, it may be more difficult to observe with sentences; this effect may be stronger when perceiving subtle speech cues in phonetics-based tasks such as identification of syllables (Zuk et al. 2013) or words (in the present study), but this effect may diminish for linguistically rich materials, such as sentences, where listeners can compensate degradations using linguistic skills as well (Benard et al. 2014). Hence, overall, the present data combined with past studies imply that there could be some transfer of music training to better perception of speech, especially in degraded listening conditions, but this effect seems to be rather small. Further, this is perhaps a consequence of the musician advantage being mainly due to better processing of low-level acoustic cues, instead of a better overall cognitive processing.

The musician effect may be stronger in speech-related tasks in which pitch cues are more important. After all, perception of speech prosody is vital to real-life speech communication and depends strongly on perception of pitch cues (Wennerstrom 2001; Besson, Chobert, and Marie 2011). One novel aspect of the present study was to include the emotion identification task to explore this idea (Exp. 2). In this test, musicians were expected to have an advantage due to better utilization of pitch cues, as in comparison to neutral speech, angry and happy speech exhibit a wider pitch range as well as a higher mean pitch, while sad speech has a narrower range and lower mean pitch (Luo et al. 2007; Banse and Scherer 1996). In line with this idea, Globerson et al. (2013) had observed that listeners with better F0 identification also exhibited better emotion identification in speech. However, other acoustic cues also contribute to vocal emotion identification, such as the level and the range of the duration and amplitude (controlled for in the present study), but also vocal energy, tempo, and pausing (not controlled; Hubbard and Assmann 2013); hence, it was not known before the present study if musician advantage indeed would also present an advantage in perception of vocal emotion in speech. In the present study, we measured emotion identification in a nonsense word (thereby removing any semantic cues) in two versions; once with all cues intact, and once with normalized duration and amplitude cues, leaving mainly the pitch cues intact. There was a small but significant overall group effect, with no interactions with presence or absence of CI simulations or of normalization of other cues than pitch, confirming that generally musicians perceived vocal emotion in speech better than non-musicians. Consistent with previous literature (Thompson, Schellenberg, and Husain 2004; Besson et al. 2007), the present data suggest that musicians may better utilize the pitch cues for vocal emotion identification, but interestingly, this is a persistent effect as they do so even when pitch cues are degraded through a CI simulation.

Note that, although twenty-five musicians and non-musicians were recruited based on a power-analysis prior to the study, the observed power for some analyses was low. This could either mean that there were not enough participants and/or that the musician effect was too small. For example, the observed power for the sentence test in stationary noise was 0.38 (Table 3). A power analysis based on the present results indicated that there would need to be a very large number of participants to achieve adequate power. Therefore, a musician effect for this specific test would not likely be found by increasing the number of participants in a realistic manner, and such a small effect might not be relevant in daily life.

On the other hand, the observed power for the emotion test was 0.56 (Table 4), and while low, this was sufficient to produce statistically significant effects. For this test, to achieve power = 0.80, the number of participants would need to be increased to 46. As such, for this test, further research with more participants has the potential to produce more significant differences between musicians and non-musicians.

Lastly, in this study, musicians were defined in terms of musical training experience, with the presumption that the musician effect was due to greater time spent with music training (e.g., Micheyl et al. 2006; Hyde et al, 2009), and not due to a genetic component or intrinsic talent (e.g., Peretz 2006). In the present study, musicians and non-musicians were tested using a wide range of listening tasks. Had the musicians shown a similar overall advantage in all tasks, it would be difficult to ascertain whether the musician effect was due to musical training per se or to musicians' genetic disposal to better listen to and process all sounds (speech or music). Because the musician effect varied across listening tasks, the present data supports the effect of musical training more than of genetic disposition, supporting our initial assumption. Previous studies with children also support this position. Children randomly assigned to a musical training group performed better in pitch and speech prosody perception and exhibited enhanced linguistic skills than children who were not given such training (Thompson, Schellenberg, and Husain 2004; Hyde et al. 2009; Moreno et al. 2009). Musical training has also been shown to enhance pediatric CI users' pitch perception (Chen et al. 2010). These studies, combined with the present data, suggest that the musician effect indeed seems to be a result of extensive musical training (Barrett et al. 2013).

Effect of the CI simulation

For all test conditions, mean performance was poorer with the CI simulation than with unprocessed speech, for both musicians and non-musicians. The effect of the CI simulation was more pronounced for more difficult listening tasks (e.g., speech recognition in noise, MCI). The musician effect persisted (or appeared, in the case of speech perception) with the application of the CI simulation, hinting that musicians were better able to extract acoustic cues in degraded conditions than non-musicians.

Interestingly, the effect of different types of noise also varied between unprocessed and CI-simulated conditions. In NH, a release of masking is observed when same listeners are tested with a steady noise vs. a fluctuating noise, usually resulting in better speech perception performance with the latter (Miller and Licklider, 1950, Başkent et al, 2014). This improvement is usually attributed to the glimpses of speech available through the valleys, i.e., low-level portions of the fluctuating noise, which provide samples of the speech that the listener can make use of to restore speech for enhanced intelligibility. In the present study, while there was such release from masking for unprocessed speech with fluctuating maskers, performance worsened with fluctuating maskers for the CI simulation. Such effects of dynamic maskers have been previously observed with real CI users and in CI simulations (Nelson et al. 2003; Fu and Nogaki 2005). The limited spectral resolution, due to both the limited number of channels and to interactions between channels, is thought to increase susceptibility to fluctuating maskers in both CI users and CI simulations. Further, recent work by Bhargava et al. (2014) showed that perhaps the reduced quality of the speech glimpses due to signal degradations in CIs make them also more difficult to utilize for the top-down reconstruction of speech in fluctuating noise. These factors can also limit melodic pitch perception in CI simulations. For example, Crew et al. (2012) showed that, even when the number of channels was increased, MCI performance was quite poor when there was substantial channel interaction in the CI simulations. Most likely, the current spread across electrodes in real CIs similarly causes spectral smearing, reducing the functional spectral resolution to be less than the number of nominal channels, thereby limiting the release from masking, as well as pitch perception.

Note that sinewave vocoding was used for the present CI simulation, rather than noise-band vocoding. The sinewave vocoder was used because of the greater specificity in terms of place of cochlear stimulation, as well as better representation of the temporal envelope, which may be "noisier" with noise-band carriers (e.g., Fu et al. 2005). One potential problem with sinewave vocoding, however, is the introduction of side-bands around the carrier frequency. Such side-band information would not be available in the case of real CIs. Although these side-bands may have provided additional (albeit weak) spectral cues beyond the 8 sinewave carriers, these cues would have been available to both musicians and non-musicians in this study. It may be that musicians were better able to use this side-band information, or were better able to use pitch cues encoded in the temporal envelope. Either way, musicians in general performed better than non-musicians in the CI simulation. This observation gives support to previous literature (Gfeller et al. 2000; Galvin, Fu, and Nogaki 2007; Galvin et al. 2012; Looi, Gfeller, and Driscoll 2012), and implies that musically trained CI users might be better able to perceive much-weakened pitch cues delivered by their devices (e.g., Fuller et al. 2014a; 2014b).

Implications for cochlear implant users

The patterns of musician effect observed with unprocessed stimuli did not change largely with the CI simulations, except for generally poorer performance, and in case of speech intelligibility, the musician effect only appeared after the CI simulation was applied. This implies that the musician effect seems to persist despite the signal degradations associated with CI signal processing, or may become even more important in the presence of such degradations where listeners can benefit even more greatly if they can perceive any acoustic cues, albeit weak. While this sounds promising, one has to be cautious before drawing strong conclusions regarding actual CI users, whose demographics vary from that of young NH populations, and who also have to deal with additional factors related to the device front-end processing and nerve-electrode interface. One important consideration is that most post-lingually deafened CI users are typically older than the present study participants (Blamey et al. 2013), and have experienced a period of auditory deprivation (Lazard et al. 2012). Age alone can alter the cognitive and linguistic processes needed for speech perception in noise (e.g., Başkent et al. 2014), and auditory deprivation may lead to structural changes in the brain, affecting overall sound perception (e.g., Lazard et al. 2014). Thus, the sometimes small musician effects in this study, measured under ideal and well-controlled conditions, may be even smaller in actual CI users. Alternatively, to their benefit, real CI users will have had much greater experience with the CI signal processing than the NH participants of the present study had experience with the simulated CI. As the actual users of CIs have to rely on these degraded signals exclusively, and will have (had) more time to practice with them, the small effects observed in this study may have greater consequences for actual CI users' real-life performance.

Previous studies have shown significant benefits of musical training after implantation for post-lingually deafened CI users' music perception (Gfeller et al. 2000; Gfeller et al. 2002; Galvin, Fu, and Nogaki 2007; Driscoll et al. 2009; Galvin et al. 2012). In the present study, musical training, the main factor that differentiated the musician group from the non-musician group, was associated with better performance as pitch cues became more important in the listening task. Training melodic pitch perception in CI users may also benefit music perception and speech perception where pitch cues are relevant (emotion recognition, prosody perception, segregation of speech from background noise or distractor signals, etc.). However, such training will likely differ from the long-term music training experienced by the present group of NH musicians. Learning to play an instrument, with spectro-temporal fine-structure cues available and over a period of many years, may give rise to robust central pitch representations. Training melodic pitch perception after implantation may not provide such robust patterns. On the other hand, an earlier training provided to hearing-impaired children before they reach the level of profound hearing loss may provide positive results, due to yet strong plasticity experienced in childhood (Hyde et al, 2009; Moreno et al, 2009; Yucel et al, 2009; Torppa et al, 2014). Further research with pre- and post-lingually deafened CI musicians and non-musicians, with or without music training provided, may reveal whether patterns developed during previous acoustic hearing or during post-implantation electric hearing may benefit pitch, music, and speech perception after implantation.

CONCLUSIONS

In this study, performance of musicians and non-musicians was compared for a variety of speech and music listening tasks, with and without the spectro-temporal degradations associated with CI signal processing. Major findings include:

1. Cross-domain (music training to speech perception) effects were weak for speech intelligibility. The musician effect was minimal for word identification in noise, and non-existent for sentence identification in noise.
2. As pitch cues became more important for the listening task (i.e., vocal emotion identification or melodic contour identification), the musician effect was more pronounced, suggesting that the musician effect may be rooted in better pitch perception.
3. Musicians tended to outperform non-musicians when listening to the CI simulation, especially for the melodic contour identification task. This suggests that musicians were better able to extract the relatively weak pitch and timbre information encoded in the CI simulations.

REFERENCES

- Banse, R, and Scherer, KR, 1996. Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology 70, 614-36.
- Başkent, D, and Shannon, RV, 2006. Frequency transposition around dead regions simulated with a noiseband vocoder. The Journal of the Acoustical Society of America 119, 1156.
- Başkent, D, and Shannon, RV, 2007. Combined effects of frequency compression-expansion and shift on speech recognition. Ear and Hearing 28, 277-89.
- Başkent, D, van Engelshoven, S, and Galvin III, JJ, 2014. Susceptibility to interference by music and speech maskers in middle-aged adults. J Acoust Soc Am 135, EL147-EL153.
- Barrett KC, Ashley R, Strait DL, Kraus N, 2013. Art and science: how musical training shapes the brain. Front Psychol.16;4:713.
- Bhargava, P, Gaudrain, E, and Başkent, D, 2014. Top-down restoration of speech in cochlear-implant users. Hear Res 309, 113-123.
- Benard, MR, Mensink, JS, Başkent, D, 2014. Individual differences in top-down restoration of interrupted speech: Links to linguistic and cognitive abilities. The Journal of the Acoustical Society of America 135, EL88-EL94.
- Besson, M, Schon, D, Moreno, S, Santos, A, and Magne, C, 2007. Influence of musical expertise and musical training on pitch processing in music and language. Restorative Neurology and Neuroscience 25 (3-4): 399-410.
- Besson, M, Chobert, J, and Marie, C, 2011. Transfer of training between music and speech: Common processing, attention, and memory. Frontiers in Psychology 2, 94.
- Bialystok, E, and DePape, AM, 2009. Musical expertise, bilingualism, and executive functioning. Journal of Experimental Psychology: Human Perception and Performance 35, 565–574.
- Blamey, P, Artieres, F, Başkent, D, et al., 2013. Factors affecting auditory performance of post-linguistically deaf adults using cochlear implants: an update with 2251 patients. Audiol Neurotol. 18,36-47,
- Boersma, P, and Weenink, D, 2012. PRAAT: Doing phonetics by computer (version 5.3.16). Institute of Phonetic Sciences, University of Amsterdam. Available at: http://www.fon.hum.uva.nl/praat.
- Bosman, AJ, and Smoorenburg, GF 1995. Intelligibility of Dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. Audiology 34, 260-84.
- Chartrand, JP, and Belin, P, 2006. Superior voice timbre processing in musicians. Neuroscience Letters 405, 164-7.
- Chen, JK, Chuang, AY, McMahon, C, Hsieh, JC, Tung, TH, and Li, LP, 2010. Music training improves pitch perception in prelingually deafened children with cochlear implants. Pediatrics 125, e793-800.
- Crew, JD, Galvin, JJ 3rd, Fu, QJ, 2012. Channel interaction limits melodic pitch perception in simulated cochlear implants. The Journal of the Acoustical Society of America 132, EL429-35.
- Deguchi, C, Boureux, M, Sarlo, M, Besson, M, Grassi, M, Schon, D, and Colombo, L, 2012. Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. Brain Research 1455, 75-89.
- Driscoll, VD, Oleson, J, Jiang, D, and Gfeller, K, 2009. Effects of training on recognition of musical instruments presented through cochlear implant simulations. Journal of the American Academy of Audiology 20, 71-82.
- Friesen, LM, Shannon, RV, Başkent, D, and Wang, X, 2001. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. The Journal of the Acoustical Society of America 110, 1150-63.
- Fu, Q. J., S. Chinchilla, and J. J. Galvin, 3rd. 2004. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. Journal of the Association for Research in Otolaryngology: JARO 5, 253-60.

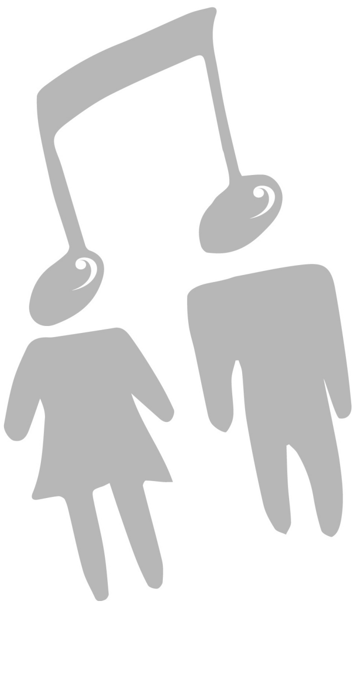
- Fu, QJ, Nogaki G, 2005, Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. Journal of the Association for Research in Otolaryngology: JARO 6, 19-27.
- Fuller, C, Free, R, Maat, B, and Başkent, D, 2012. Musical background not associated with self-perceived hearing performance or speech perception in postlingual cochlear-implant users. The Journal of the Acoustical Society of America 132, 1009-1016.
- Fuller, C, Galvin III, JJ, Free, R, and Başkent, D, 2014a. Musician effect in cochlear implant simulated gender categorization. The Journal of the Acoustical Society of America 135, EL159-65.
- Fuller, C, Gaudrain, E, Clarke, J, Galvin III, JJ, Fu, QJ, Free, R, Başkent, D, 2014b. Gender categorization is abnormal in cochlear-implant users. Journal of the Association for Research in Otolaryngology: JARO, revision submitted.
- Galvin, JJ,3rd, Fu, QJ and Nogaki, G, 2007. Melodic contour identification by cochlear implant listeners. Ear and Hearing 28, 302-19.
- Galvin, JJ,3rd, Fu, QJ and Oba,S, 2008. Effect of instrument timbre on melodic contour identification by cochlear implant users. The Journal of the Acoustical Society of America 124, EL189-95.
- Galvin JJ, 3rd, Fu, QJ, and Shannon, RV, 2009. Melodic contour identification and music perception by cochlear implant users. Annals of the New York Academy of Sciences 1169, 518-33.
- Galvin JJ, 3rd, Eskridge, E, Oba, S, and Fu, QJ, 2012. Melodic contour identification training in cochlear implant users with and without a competing instrument. Seminars in Hearing 33, 399-409.
- Gfeller, K, Witt, S, Stordahl, J, Mehr, M, and Woodworth, G, 2000. The effects of training on melody recognition and appraisal by adult cochlear implant recipients. Journal of the Academy of Rehabilitative Audiology: JARA 33, 115-138.
- Gfeller, K, Witt, S, Adamek, M, Mehr, M, Rogers, J, Stordahl, J, and Ringgenberg, S, 2002. Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. Journal of the American Academy of Audiology 13, 132-45.
- Gfeller, K, Turner, C, Mehr, M, Woodworth, G, Fearn, R, Knutson, JF, Witt, S, and Stordahl, J, 2002. Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. Cochlear Implants International 3, 29-53.
- Gfeller, K, Turner, C, Oleson, J, Zhang, X, Gantz, B, Froman, R, and Olszewski, C, 2007. Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. Ear and Hearing 28, 412-23.
- Globerson, E, Amir, N, Golan, O, Kishon-Rabin, L, and Lavidor, M, 2013. Psychoacoustic abilities as predictors of vocal emotion recognition. Attention, Perception, and Psychophysics 78, 1799-1810.
- Goudbeek, M, and Broersma, M, 2010. Language specific effects of emotion on phoneme duration. Paper presented at Eleventh Annual Conference of the International Speech Communication Association.
- Greenwood, DD, 1990. A cochlear frequency-position function for several species—29 years later. The Journal of the Acoustical Society of America 87, 2592-605, 13.
- Heng J, Cantarero G, Elhilali M, Limb CJ, 2011. Impaired perception of temporal fine structure and musical timbre in cochlear implant users. Hearing Research 2011 280(1-2):192-200.
- Hubbard, DJ, and Assmann, PF, 2013. Perceptual adaptation to gender and expressive properties in speech: The role of fundamental frequency. The Journal of the Acoustical Society of America 133, 2367-76.
- Hyde, KL, Lerch, J, Norton, A, Forgeard, M, Winner, E, Evans, AC, and Schlaug, G, 2009. The effects of musical training on structural brain development: A longitudinal study. Annals of the New York Academy of Sciences 1169, 182-6.
- Kong YY, Cruz R, Jones JA and Zeng FG, 2004. Music perception with temporal cues in acoustic and electric hearing. Ear and Hearing 25, 173-185.
- Kraus, N, and Chandrasekaran, B, 2010. Music training for the development of auditory skills. Nature Reviews Neuroscience 11, 599-605.

- Lazard, DS, Vincent, C, Venail, F, et al. 2012. Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: A new conceptual model over time. PloS One 7, e48739.
- Lazard DS, Innes-Brown H, Barone P, 2014. Adaptation of the communicative brain to post-lingual deafness. Evidence from functional imaging. Hear Research 307:136-43.
- Limb, CJ, and Roy, AT, 2014. Technological, biological, and acoustical constraints to music perception in cochlear implant users. Hearing Research 308:13-26.
- Looi, V, Gfeller, K and Driscoll, V 2012. Music appreciation and training for cochlear implant recipients: A review. Seminars in Hearing 33, 307-34.
- Luo, X, Fu, QJ and Galvin JJ, 3rd, 2007. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. Trends in Amplification 11 (4): 301-15.
- McDermott, HJ, 2004. Music perception with cochlear implants: A review. Trends in Amplification 8 (2): 49-82.
- Micheyl, C, Delhommeau, K, Perrot, X, and Oxenham, AJ, 2006. Influence of musical and psychoacoustical training on pitch discrimination. Hearing Research 219, 36-47.
- Miller, G.A., Licklider, J.C.R., 1950. The intelligibility of interrupted speech. J. Acoust. Soc. Am. 22, 167.
- Moreno, S, Marques, C, Santos, A, Santos, M, Castro, SL, Besson, M, 2009. Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. Cereb Cortex 19(3):712-23.
- Moreno, S, Bialystok, E, Barac, R, Schellenberg, EG, Cepeda, NJ, and Chau, T, 2011. Short-term music training enhances verbal intelligence and executive function. Psychological Science 22, 1425–1433.
- Musacchia, G, Strait, D, and Kraus, N 2008. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hearing Research 241, 34-42.
- Nelson, PB, Jin, SH, Carney, AE, Nelson DA, 2003. Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. The Journal of the Acoustical Society of America 113, 961-968.
- Oxenham, AJ, 2008. Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. Trends in Amplification 12, 316-31.
- Parbery-Clark, A, Skoe, E, Lam, C, and Kraus, N, 2009. Musician enhancement for speech-in-noise. Ear and Hearing 30, 653-61.
- Peretz, I, 2006. The nature of music from a biological perspective. Cognition 100, 1-32.
- Plomp, R, and Mimpen, AM, 1979. Speech-reception threshold for sentences as a function of age and noise level. The Journal of the Acoustical Society of America 66, 1333-42.
- Ruggles, DR, Freyman, RL, and Oxenham, AJ, 2014. Influence of musical training on understanding voiced and whispered speech in noise. PLoS ONE 9(1):e86980.
- Schon, D, Magne, C, and Besson, M, 2004. The music of speech: Music training facilitates pitch processing in both music and language. Psychophysiology 41, 341-9.
- Shannon, RV, Zeng, FG, Kamath, V, Wygonski, J, and Ekelid, M, 1995. Speech recognition with primarily temporal cues. Science 270, 303-4.
- Stickney, GS, Zeng, FG, Litovsky, R, and Assmann, P, 2004. Cochlear implant speech recognition with speech maskers. The Journal of the Acoustical Society of America 116, 1081.
- Strait, DL, Kraus, N, Skoe, E, and Ashley, R, 2009. Musical experience and neural efficiency: Effects of training on subcortical processing of vocal expressions of emotion. The European Journal of Neuroscience 29, 661-8.
- Thompson, WF, Schellenberg, EG, and Husain, G, 2004. Decoding speech prosody: Do music lessons help? Emotion 4, 46-64.
- Torppa R, Faulkner A, Huotilainen M, Järvikivi J, Lipsanen J, Laasonen M, Vainio M, 2014. The perception of prosody and associated auditory cues in early-implanted children: the role of auditory working memory and musical activities. Int J Audiol 53(3):182-91.

- Versfeld, N. J., L. Daalder, J. M. Festen, and T. Houtgast. 2000. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. The Journal of the Acoustical Society of America 107, 1671-84.
- Wennerstrom, A, 2001. The music of everyday speech: Prosody and discourse analysis. Oxford University Press. New York: Oxford University Press.
- Wilson, B. S., C. C. Finley, D.T. Lawson et al. 1991. Better speech recognition with cochlear implants. Nature 352:236–238.
- Won, JH, Drennan, WR, Kang, RS, and Rubinstein, JT, 2010. Psychoacoustic abilities associated with music perception in cochlear implant users. Ear and Hearing 31, 796-805.
- Wong, PC, Skoe, E, Russo, NM, Dees, T, and Kraus, N, 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nature Neuroscience 10, 420-2.
- Yang, LP, and Fu, QJ, 2005. Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. The Journal of the Acoustical Society of America 117, 1001-4.
- Yucel E, Sennaroglu G, Belgin E, 2009. The family oriented musical training for children with cochlear implants: speech and musical perception results of two year follow-up. Int J Pediatr Otorhinolaryngol 73(7):1043-52.
- Zuk, J, Ozernov-Palchik, O, Kim, H, Lakshminarayanan, K, Gabrieli, JDE, Tallal, P, and Gaab, N, 2013. Enhanced syllable discrimination thresholds in musicians. Plos One 8, e80546.

Part three:

The effect of music therapy and training in cochlear-implant users

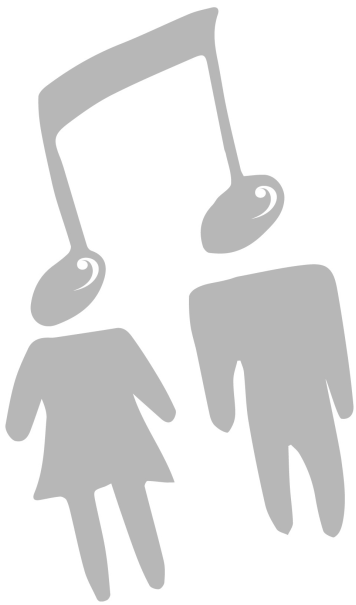


Chapter 9

The effect of musical training and music therapy on speech and music perception in cochlear-implant users

Christina D. Fuller[1,2], John J. Galvin III [1,2,3,4], Bert Maat[1,2], Deniz Başkent [1,2], Rolien H. Free[1,2]

to be submitted



[1] University of Groningen, University Medical Center Groningen, Department of Otorhinolaryngology / Head and Neck Surgery, Groningen, The Netherlands

[2] University of Groningen, Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, Groningen, The Netherlands

[3] House Research Institute, Division of Communication and Auditory Neuroscience, Los Angeles, CA, USA

[4] Department of Head and Neck Surgery, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA

ABSTRACT

Music is reported to be the second most important acoustical stimulus by cochlear-implant (CI) users. Yet music is not well-perceived, nor enjoyed by CI users. Previously, normal-hearing musicians have been shown to have better speech perception in noise, pitch perception in speech and music, as well as a better working memory and enhanced neural encoding of speech, compared to non-musicians. Based on these findings, we hypothesize that music therapy and musical training may have a positive effect on speech and music perception in CI users.

Three groups of CI users were recruited for six weeks of musical training, music therapy and non-musical training for 2 hours per week. These different types of training were selected to vary in their level of experimental control and human interaction. Musical training involved individual computerized training with melodic contour identification and instrument recognition. Music therapy involved group exercises involving rhythm, musical speech, singing, emotion identification and improvisation. Non-musical training involved group activities involving writing, cooking, and woodworking. Before and after the training, all participants were tested behaviorally for speech and music-related tasks (emotion identification, melodic contour identification). Quality of life was quantified using the Nijmegen Cochlear Implant Questionnaire.

In general, training effects were observed within domain (from musical training to better melodic contour identification), with little transfer across domains (no effect of any of the three training approaches on speech perception, but an effect of music therapy on emotion identification). The music therapy group also reported enhanced subjective perceptual skills. None of the training methods showed an effect on quality of life.

Given the short duration of training, it is promising that music therapy already showed a cross-domain positive effect on emotion identification, along with subjective reports of benefit. It is possible that the interactive nature of music therapy was useful; however, further research is needed with more participants and with longer durations of training.

INTRODUCTION

Recent research has shown that normal-hearing (NH) musicians have certain benefits for the perception of auditory signals. For example, musicians have been shown to have a better perception of pitch and a better ability to detect pitch changes in foreign languages (Marques et al. 2007; Besson et al. 2007). This indicates that musical training and thus being a musician can create benefits for music related auditory perception. Moreover, musicians, even though the results of the studies are more ambivalent, have also been shown to have a better perception of speech, both in quiet and in noise (Ruggles, Freyman, and Oxenham 2014; Parbery-Clark et al. 2009; Boebinger et al. 2015; Swaminathan et al. 2015; Zendel and Alain 2013). This indicates that musical training could also create a benefit for speech perception, a possible positive transfer or training effect of music on speech.

These advantages in auditory perception from musical training make musicians into a very interesting research group, as they can serve as a model for neural plasticity of auditory perception (Herholz and Zatorre 2012). Furthermore, the neural plasticity of musical training could be interesting for other groups of listeners, so not only for normal hearing (NH) listeners, but also for hearing deprived persons, such as cochlear implant (CI) users (Fuller et al. 2014b; Fuller et al. 2014a). This study will focus on the possible positive effects of music related training on auditory perception and the quality of life in CI users.

Music in CI users

Music is the second most important auditory signal for CI users, but yet not well perceived or enjoyed (Drennan and Rubinstein 2008; Gfeller et al. 2000; Philips et al. 2012; Fuller et al. under revision). The perception of music and its basic elements (rhythm, pitch, melody and timbre) is less exact in CI users compared to NH listeners. This difference in perception is partially due to the difference between acoustic and electric hearing, thus factors related to the CI itself. Due to the limited amount of simulation sites in the cochlea the perceived tonotopy is imprecise. Furthermore, spectral and temporal limitations of the device caused by the speech processing strategies that only retain the slow temporal envelope cause the perception of three of four of the basic elements of music, pitch, melody and timbre to be perceived poorer in CI users in comparison to NH listeners (NH) (see for a review: McDermott 2004; Galvin, Fu, and Nogaki 2007; Gfeller et al. 2007; Gfeller et al. 2002; Kong et al. 2009; Looi and She 2010; Looi, Gfeller, and Driscoll 2012; Limb and Roy 2014). Only rhythm is perceived with almost similar accuracy in CI users as in NH listeners (Gfeller et al. 2007; Kong et al. 2004).

Next to the limitations caused by the device – implant-related limitations – CI users have suffered from changes in the auditory nervous system, posing patient-related limitations as well (Limb and Roy (2014) for a review). For example, most post-lingually deafened implantees have a possibly deprived peripheral and/or central auditory pathway, caused by different etiologies, different survival of spiral ganglia or caused by changes in cognitive elements, such as can develop after a long period of profound deafness. The patient- and implant-related factors together cause a degraded perception of the elements of music (Başkent and Gaudrain 2016). These factors not only cause difficulties in music perception, but also affect the perception of pitch. A deprived perception of pitch causes difficulties in areas such as speech-on-speech perception, the identification of vocal emotions or the gender of a talker, as well as on the perception of more complex and more pitch-related signals such as melodies and the perception and appreciation of music itself (Philips et al. 2012; McDermott 2004; Galvin, Fu, and Nogaki 2007; Gfeller et al. 2007; Looi, Gfeller, and Driscoll 2012; Wright and Uchanski 2012; Gilbers et al. 2015; Fuller et al. 2014c; Xin, Fu, and Galvin 2007).

Research using musical training could therefore serve several purposes in CI users: first, it could give us insight in the perception of music in CI users; second, it could be a fun way to train and enhance the music and possibly speech perception; and third, it could serve as a model for the enhancement of the neural plasticity of CI users for auditory signals.

Musician effect in CI

Because of the limitations of both the device and the patient for music perception, it is unclear if the 'musician effect', the possible positive effect of musical training on speech perception, would persist and/or exist in CI users. The effect of musical training or rehabilitation has been less well explored in CI users. Thus far, focused music training in CI users shows that melodic contour processing, timbre recognition and complex melody tasks recognition can be improved (Oba, Fu, and Galvin 2011; Gfeller et al. 2002; Looi, Gfeller, and Driscoll 2012). All these studies focus on the within domain neural plasticity, i.e. if you train music this only affects the perception of music itself. To see whether the musician effect also exists in auditory deprived situations such as with electrical hearing in CI users, Fuller et al. (2014b) recently studied NH musicians and non-musicians listening to CI simulated stimuli, both speech and music. The results showed a small, transfer effect of musical training for word identification in noise. No effect for the perception of sentences both in quiet and in noise was found. A musician effect, however, was shown for the identification of emotions, as well as for the identification of melodic contours. It was suggested that the musician effect in this study could be based on a better perception of pitch. Still the question remains whether the musician effect persists in CI users. Thus far only two studies investigated the effect of musical training on speech perception in CI users. A pilot study with two CI users showed a small effect of purely instrumental, melodic contour identification training on speech in noise reception thresholds in one CI user and an improvement of prosody in words perception in the other CI user (Patel 2014). A second study by Lo et al. (2015) showed an effect of two melodic contour training programs on speech perception in 16 CI users and 12

NH listeners. One group was trained with different semitone interval sizes, the other with the duration of the notes. Both groups were trained for six weeks and tested before and after for sentences in four talker babble noise perception, consonant discrimination and prosody identification. Results indicated a small effect of both training methods in CI users on consonant identification and prosody identification, no effect on sentences in babble was shown. The results from all these studies indicate possibilities for cross-domain training effect of music on speech in CI users. Nevertheless all these studies lack a control group to see whether no intervention could give a better perception as well.

Methods of training

Conventional auditory training seems to be effective in CI users. Bottom up auditory training has been shown to improve the auditory perception of the stimulus trained with, for example improvement of speech in noise recognition by training speech in noise (Fu and Galvin 2008; Ingvalson et al. 2013). Which method is best for training music perception and enjoyment, and investigating the transfer effect on speech is unclear.

In this study three different types of training and groups will be exploited. First, an individualized, computerized musical training using melodic contour identification; second, a group-wise music therapy and third, a control group that receives group-wise training that is not related to music or auditory perception.

The first group, the musical training group received training based on an individual, computerized training. Computer based training, which allows for an easy, individual training for large numbers of trials, has been used in different CI training studies, regarding different topics of perception, including as described above music perception (see: Galvin, Fu, and Nogaki 2007; Başkent et al. 2016; Benard and Başkent 2013; Fu, Chinchilla, and Galvin 2004; Fu, Nogaki, and Galvin III 2005; Galvin et al. 2012; Stacey et al. 2010; Loebach and Pisoni 2008; Nogaki, Fu, and Galvin 2007; Stacey and Summerfield 2007; Stacey and Summerfield 2008). The ideal is a limited number of stimuli and simple tasks (Oba, Fu, and Galvin 2011). In this study we chose the melodic contour identification task as the stimulus (Galvin et al., 2007), as was used in Patel et al. (2014) and Lo et al. (2015).

Next to an individualized and computerized training method, a second group will be trained using music therapy. Thus far, to the best of our knowledge, studies using music therapy show a positive effect on QoL and cognition in patient groups with for example dementia and Parkinson, indicating the possibility of a positive effect of music therapy on QoL in CI users (Van de Winckel et al. 2004; Pacchetti et al. 2000). Recently a pilot study by Hütter et al. (2015) in adult CI users was performed using an individualized music therapy program of 10 sessions of 50 minutes, specifically addressed to the individual needs of the CI user. The program was focused on the perception of musical parameters, prosody and complex acoustic situations and started shortly after the initial activation of the speech processor

(Hütter et al. 2015). The preliminary results suggest improvements in subjective music and overall hearing perception. Music therapy consists of training both with speech, music and motoric training by playing the instrument. It has been suggested that exploiting the effects of multimodality by actively playing a musical instrument might be more beneficial for/in creating neuroplasticity, causing the transfer effect perhaps to be more apparent with this type of training (Herholz and Zatorre 2012).

To the best of our knowledge no study has looked into the (possible translational) effect of group-wise music therapy on the perception of music and speech in CI users. By using these training methods in this study we were interested to see if music training or therapy makes a difference on the auditory performance, but also to see whether the improvement of enjoyment of music or the potential of regular meetings with a group of CI users, without any musical influences, might have an effect on the health-related quality of life (QoL). Therefore as a control for a possible group effect or just training effect on QoL and perhaps on the perception of speech and music, a third, control group that only comes together will be incorporated. The third group will not exploit any musical activities.

In this study the effect of a musical training or therapy program, which could be added to the current CI rehabilitation program for the improvement of music perception and enjoyment, will be investigated in a prospective design. A possible positive transfer effect of this training to improvements in speech or music related auditory performance and/or to the quality of life of CI users is investigated.

MATERIALS AND METHODS

Participants

The current study is a prospective study with three training groups that consisted of CI users: 1) the musical training group; 2) the music therapy group; 3) the non-musical training group: the control group

Nineteen post-lingually deafened, adult CI users were recruited via the University Medical Center Groningen (UMCG; see Table 1 for more details). All participants were native Dutch speakers, had a CI for longer than one year, and had no neurological disorders. One of the CI users was a bilateral CI user. Four CI users were bimodal users. Before the study started, written and oral information about the protocol was provided, and informed consent was obtained from all participants. The travel costs and the testing time were financially reimbursed in accordance with the department policy.

Procedures

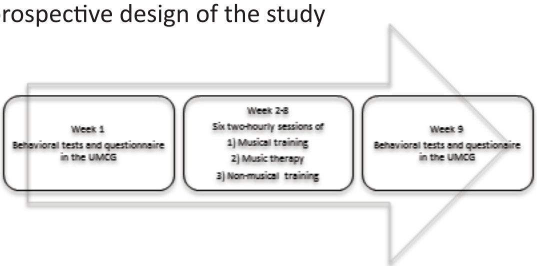
The prospective design of the study in time is depicted in Flowchart 1. Before the training, all participants were first tested with the baseline tests in week 1. These tests constituted of behavioral tests (word, speech, gender and emotion identification and melodic contour

TABLE 1. Demographic characteristics of the CI users.

	Musical Training	Music Therapy	Non-musical Training
Gender (M:F)	3:3	3:4	5:1
Age range (Mean (yrs))	70-78 (73)	56-71 (64)	65-80 (72)
Brand CI			
Cochlear	5	5	3
CI24R CS		1	
CI24RE CA	2	2	3
CI24R CA	1	1	
Ci24R k	1		
CI512	1	1	
Advanced Bionics	1	2	3 (1 bilateral user)
HiRes 90K Helix	1	2	3
Etiology			
Unknown	5	7	3
Sudden deafness			1
Trauma			1
Progressive hearing loss	1		1
Bimodal	2	1	1
Years of CI use (range (yrs))	4,8 (1-11)	4.14 (1-9)	3.0 (1-5)

identification) and a quality of life questionnaire. After the first set of baseline tests in week 1 the CI users were randomly distributed between the three training groups. Due to the small number of participants, no matching was attempted between the three groups. The music therapy group had seven participants; the musical training and non-musical training groups both had six participants. The training sessions were completed within weeks 2-8, and the last week (week 9) constituted the same baseline tests from week 1.

Flowchart 1. The prospective design of the study



9

Behavioral tests

The behavioral tests before (week 1) and after (week 9) training were conducted in an anechoic chamber at UMCG. Total testing time was two hours including the filling of the questionnaire. All CI users used their own CI(s), with no HA in the case of bimodal users, during testing. The participants were asked to put the CI on their daily life settings and to not change these settings during testing. All CI users were seated facing a touch screen (A1 AOD 1908, GPEG International, Woolwich, UK) and a speaker at a 1-meter distance (Tannoy precision 8D; Tannoy Ltd., North Lanarkshire, UK). Stimuli were presented using iStar (http://tigerspeech.com/istar/), for the words and sentences, and Angelsound™ (Emily Shannon Fu Foundation, http://www.angelsound.tigerspeech.com/), for non-speech tests – i.e. emotion identification and melodic contour identification (MCI). All stimuli were played via a Windows computer with an Asus Virtuoso Audio Device soundcard (ASUSTeK Computer Inc. Fremont. USA). Converted to an analogue signal via a DA10 digital-to-analog converter (Lavry Engineering Inc., Washington. USA) the stimulus was played at 65 dBSPL in sound field. Except for the noises of the speech stimuli, the root mean square (RMS) intensity of all stimuli was normalized to the same value. Calibration was performed with a manikin (KEMAR, GRAS) and a sound-pressure level meter (Type 2610, Brüel Kjær and Sound & Vibration Analyser, Svan 979 from Svantek). Verbal responses on the speech tests were scored online by a student assistant in the adjacent room, as well as recorded using a DR-100 digital voice recorder (Tascam, California, USA) for offline double-check of the responses when needed.

Word identification

The first speech perception task was word identification. Stimuli included digital recordings of meaningful, monosyllabic Dutch words in CVC format [e.g. bus (‘bus’ in English), vaak (‘often’), nieuw (‘new’), etc.] taken from the clinically used NVA test (merged). Twelve lists, each of which contains twelve words spoken by a female talker, were used.

Word identification was tested in four conditions: in quiet and in steady, speech-shaped noise at three signal-to-noise ratio’s (SNRs) (+10, +5 and 0 dB). One randomly selected list out of 12 lists was used to test each condition. No list of words was repeated within a participant. The words were randomly presented. The participant was asked to repeat the word out loud as accurately as possible, and if in doubt, to guess. The software automatically calculated the percentage correct of the phonemes. Stimuli were only played once and no feedback was provided.

Sentence identification

The second speech perception task was sentence identification. Sentences used were syntactically correct Dutch sentences with a meaning and a semantic context (Plomp and

Mimpen 1979). The corpus contains digital recordings of 10 lists of 13 sentences (4 to 8 words per sentence) spoken by a female talker. Sentence identification was measured using three types of noise: 1) steady, speech-shaped noise; 2) fluctuating, speech-shaped noise; and 3) 6-talker babble (Yang and Fu 2005).

Sentence identification was measured in quiet and in noise. One list of 13 sentences was used to test each condition. Sentence lists were randomly chosen from the 10 lists in the test corpus. No list was repeated per participant per session. The participant was asked to repeat the sentence out loud as accurately as possible. The observer in the adjacent room scored the words in the sentence correctly identified. In quiet only, the performance was calculated in terms of the percentage correct of words in the test list correctly identified. For the noise conditions, the speech reception threshold (SRT), defined as the SNR needed to give a 50% correct full sentence identification, was measured using an adaptive one-up/one-down procedure (Plomp and Mimpen 1979). The sentence and noise were presented at a target SNR and the participant was asked to repeat the sentence as accurately as possible. If all words in the sentence were correctly repeated, the SNR was reduced by 2 dB; if not all words were correctly repeated, the SNR was increased by 2 dB. The average of the reversals in SNR between trials 4-13 was reported as the SRT. To target as accurately as possible the SRT within the limited number of sentences, the initial SNR was set to +2 dB for the steady noise condition, and to +6 dB for the fluctuating and babble noise. Note that the first sentence was repeated and the SNR increased until the participant repeated the entire sentence correctly.

Emotion identification

The third behavioral test was a vocal emotion identification test. Stimuli were digital recordings of a nonce word [nutohɔmsɛpikɑŋ] made by Goudbeek and Broersma (2010) and also described and used in (merged). The nonce word was originally produced by eight professional Dutch actors with eight target emotions (‘joy’, ‘pride’, ‘anger’, ‘fear’, ‘tenderness’, ‘relief’, ‘sadness’, and ‘irritation’). Based on a pilot study with three normal hearing listeners, four actors (two female, two male) and four emotions (‘joy’, ‘anger’, ‘relief’ and ‘sadness’) were chosen for formal testing. The four emotions were selected to represent all corners of the emotion matrix based on the prevalence or absence of arousal (defined as the difference between high and low arousal emotions) or valance (defined as the difference between positive and negative emotions): 1) joy (high arousal, positive valence); 2) anger (high arousal, negative valence. 3) relief (low arousal, positive valence); and 4) sadness (low arousal, negative valence). Two recordings of each emotion from each actor were used, producing a total of 32 tokens (4 actors × 4 emotions × 2 utterances).

Participants were first familiarized with the emotion task. For the familiarization session, we have used the same target emotions as the actual test, but produced by four other actors that were not used for formal testing. In both familiarization and data collection sessions, the target emotion was randomly selected from the stimulus set and presented over the loudspeaker. Subjects indicated the emotion by touching one of four response boxes on the touch screen labeled: 'anger,' 'sadness,' 'joy,' and 'relief'. During familiarization, only visual feedback was provided on the screen in case of a correct answer, showing the emotion. In case of an incorrect answer, audio-visual feedback was provided for the correct and incorrect response, showing the emotion and playing the emotion. The actual data collection was identical to familiarization, but no feedback was provided. The software automatically calculated the percent correct score.

Melodic contour identification

The fourth behavioral test was based on a music task: the identification of the contour of a melody. The melodic contour identification (MCI) task as developed by Fu and Galvin (2007) for testing the ability of CI users in identifying melodies with different contours and different semitone spacing. The MCI test consists of nine melodic contours with 5 notes, each with changes in pitch pattern: "Rising," "Flat," "Falling", "Flat-Rising," "Falling-Rising," "Rising-Flat," "Falling-Flat," "Rising-Falling," "Flat-Falling". The A3 (220 Hz) was always the lowest note in the contours. The semitones between the successive notes were 1, 2 or 3 semitones. The duration of each note in the contour was 250 millisecond (ms). The silent interval between notes was 50 ms. The target instruments used in the testing were the piano or the organ (similar to Galvin, Fu, and Oba (2008)). MCI was tested once in quiet and twice with a competing masker instrument, as in Galvin, Fu, and Oba (2009). The masker instrument was the piano with a "Flat" contour. The base pitch of the masker was either an overlapping pitch (A3 (220 Hz) or a non-overlapping pitch (A5 (880 Hz)). This resulted in a total of six conditions: 1) piano (no masker), 2) piano with the A3, 3) piano with the A5 masker and: 1) organ (no masker), 2) organ with the A3 masker, and 3) organ with the A5 masker. The masker started and stopped at the same time as the target contour.

Questionnaires

Nijmegen Cochlear Implant Questionnaire

In weeks 1 and 9 all CI users were also asked to fill a health-related quality of life questionnaire (HRQoL), the Nijmegen Cochlear Implant Questionnaire (NCIQ), a validated CI-specific HRQoL instrument (Hinderink, Krabbe, and Van Den Broek 2000). The questionnaire consisted of six different domains that included 10 statements with a 5-point response scale. The domains were: sound perception basic, sound perception advanced, speech production, social functioning, and psychological functioning. Per statement, the scores could vary between 0 and 100 (lowest and highest, respectively). The total score was calculated as the average of the 6 domains.

Subjective perceptual skills in the music therapy group

In weeks 2 - 8 the music therapy group filled an additional, non-validated questionnaire, as part of the music therapy project. A short survey was done after every therapy session. In this survey participants could judge if they felt that they improved on different auditory tasks. The CI users could score the questions from 1-10, where 1 was the poorest and 10 the best score. The questions related to the different elements of the therapy.
Did you notice an improvement for:

- Rhythm
- Musical speech
- Perception of music
- Playing music?

STATISTICS

The results of the five behavioral tests were statistically analyzed using IBM SPSS statistics version 22. Repeated measures ANOVA's were used for the behavioral tests and the questionnaire results both for between and within group results. The factors used in the ANOVA's were the three groups (musical training, music therapy and non-musical training) and the conditions per test: for word identification the factors used were the four conditions (quiet, +10, +5 and 0 dB SNR); for sentence identification the three different noises (stationary, fluctuating and babble); for emotion identification the before and after scores; for melodic contour identification per instrument the three different conditions (no masker, A3, A5). For emotion identification also a repeated measures ANOVA for the music therapy group only was used. A p-level of < 0.05 was considered significant and was corrected in case of multiple testing. In word identification and the emotion identification tests, to correct for the skewed data based on the ceiling effect, the percent correct scores were log-transferred before applying the ANOVA.

Training

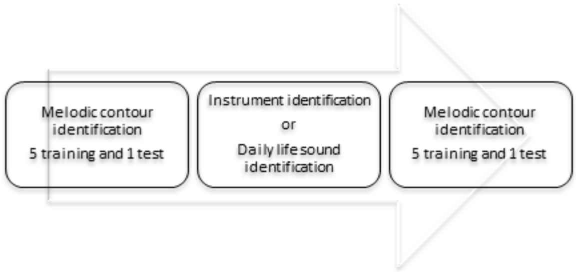
In weeks 2-8, training was provided. During the training and therapy sessions the CI users were allowed to wear their hearing aid if they were bimodal users, to make the training and therapy as comparable to reality as possible. Nevertheless during testing only the CI was used to see what the effect of the intervention was on hearing with a CI only. Furthermore the results are in this way not interfered by bimodality in the testing. The bilateral CI user was allowed to use both CI's during testing and training.

Musical training

The musical training of the study was based on MCI, and was provided via a customized computer program (Galvin, Fu, and Nogaki 2007). In the training different instruments than the ones used in testing, i.e., piano and organ, were used (violin, glockenspiel or trumpet). Stimuli were similar to the MCI described above in baseline testing, and were played using Angelsound (Emily Shannon Fu Foundation. http://angelsound.tigerspeech.com/). Each CI user was seated in a separate room, facing a personal computer and two loudspeakers. The musical training consisted of six, weekly sessions of two hours with a 15 minutes break in the middle. If needed the CI users could ask the help of the assistants that were available at every session for technical problems. The assistants set up the computers and speakers before every session. Only during the first session, the CI users would get started with the sessions by student assistants. The computer was started and the students helped the CI users with familiarization of the test software. After observing one round of training the assistants left. They were available in case of questions or computer problems. Flowchart 2 shows the procedure in time.

At the beginning of each session a written explanation of the exercises for that particular session was provided. The exercises for the melodic contour identification (MCI) went from relatively easy to difficult (Galvin, Fu, and Nogaki 2007). An easy exercise was training with a distance of six semitones between the five notes of the contour; a difficult exercise was training with a distance of only one semitone. A total of five exercises per instrument were completed. During the exercises direct audio-visual feedback was provided in case of an incorrect answer. The feedback involved playing the correct and wrong melodic contours one after another, while on the screen the melodic contour was depicted at the same time. After the five exercises the training per instrument ended with a test. During the test, all semitone distances (one to five) and all nine melodic contours were randomly presented. Furthermore, no feedback was provided. Per instrument a total of six rounds of MCI was completed per participant; five training rounds, and one test. After a full round with one instrument the participants trained with a different task: instruments identification or daily life sound identification. Instrument recognition involved recognizing different instruments by the sound only, choosing from nine different instruments. The daily life sound identification involved recognizing different daily life sounds choosing from options. Examples are recognizing a baby crying, a cat meowing or a car honking the horn. Both involved audiovisual feedback, in the same way as described for MCI. This was done to diversify the training session. After one of these in-between exercises the participants continued with the training of the MCI with a new instrument. Flowchart 2 depicts the set-up of the sessions.

Flowchart 2: The design of the musical training sessions



Music therapy

We have included music therapy as a bridge between musical training and non-musical training. Music therapy, differently than computerized music training, consists of six different types of therapy and training exercises ranging from music only to music related speech perception and to singing, as well as the motor training of playing an actual instrument. Another important difference is that it is interactive in its nature, conducted under supervision of music therapists and in interaction with both therapists and the other participants. The method used is based on practice based evidence using a bottom up approach i.e. the interaction between the therapists and the clients form the sessions. Reflection and feedback are the baseline for the changes to the first model (Migchelbrink and Brinkman 2000). For an extensive report on the development of the training in Dutch see: https://figshare.com/s/db66eb0714a5bd9496d8.

The music therapy sessions were organized under the supervision of three music therapy students and their lecturer and supervisor from the Hogeschool Utrecht, Amersfoort. All sessions were held in the activity room of the rehabilitation center of the CI team of the Northern-Netherlands and were always accompanied by the music therapy students and one of the members of the CI team.

In total the music therapy protocol consisted of six weekly sessions of two hours, with a break of 15 minutes in every session. After each therapy session, this session was evaluated by all involved individuals and, if necessary, changed in accordance to their comments. The therapy sessions were conducted in a group, mostly in a circle, with the CI users facing each other. For some exercises the therapists arranged the CI users in such a way that they had to rely only on the acoustic signal. For example, in an exercise called 'the bus': a bus driver gets in the bus (a row of chairs behind each other, respectively) with an instrument playing a rhythm; a passenger gets in the bus behind the bus driver and starts to play a different rhythm to which the bus driver has to match his or her rhythm; a second passenger enters the bus and starts to play a new rhythm to which both the bus driver and the first passenger have to match their rhythms etc. All sessions were accompanied by different musical instruments, such as a guitar, piano, drums and xylophone, played by both the music therapy students as well as the CI users. Furthermore singing and improvising was encouraged.

The music therapy involved different elements of the perception and enjoyment of music and music-related speech tasks. The elements of the sessions had a construction based on the literature of the perception of music and the elements of music with CI's (Gfeller et al. 2000; Fuller et al. under revision; McDermott 2004) from rather easy to perceive for CI users – rhythm – to difficult to perceive – improvising with music. The exercises themselves also had a gradual build up from easy rhythms with for example one instrument, to more difficult rhythms with more instruments at once. All elements were practiced via different exercises that were always first explained and shown by the therapy students. Afterwards the CI users were allowed to discuss what difficulties they encountered or what they felt.
All sessions consisted at least of the following elements, from easy to difficult:

- Rhythm
- Emotion Identification
- Musical Speech
- Singing
- Playing Music
- Improvising with music

We have included non-musical training as a control group that did not perform any musical activities, but that did need to interact with each other during different tasks. This was done to see whether also non-music related interactions made a difference for example for quality of life.

The first two sessions of two hours involved a writing course with a professional writing coach at the boat of the coach; the second pair of sessions involved a cooking course during which the participants had to collaborate to prepare different dishes at the kitchen located at the school of the deaf; the last two sessions involved a wood workshop during which the participants had to build a birdhouse under the supervision of a woodwork teacher also held at the school of the deaf. A member of the CI-team, the social worker, accompanied all sessions to explain the tasks and be available if there were any difficulties or questions. The non-musical training consisted of interactive group activities that had no connection to music. The total training consisted of a total of six weekly sessions of two hours, with a break of 15 minutes in the middle.

RESULTS

Word identification

Figure 1 shows the results for the word identification test before and after the training for the three groups in percent correct scores. The figure shows that there are no notable differences for word identification between before and after the training, or between the different training groups. The repeated measures ANOVA used all four conditions (quiet, +10, +5 and 0 dB SNR) and all three training groups (musical training, music therapy, non-musical training) as factors. No significant effect of training for the word identification within or between the training groups between before and after the sessions was shown.

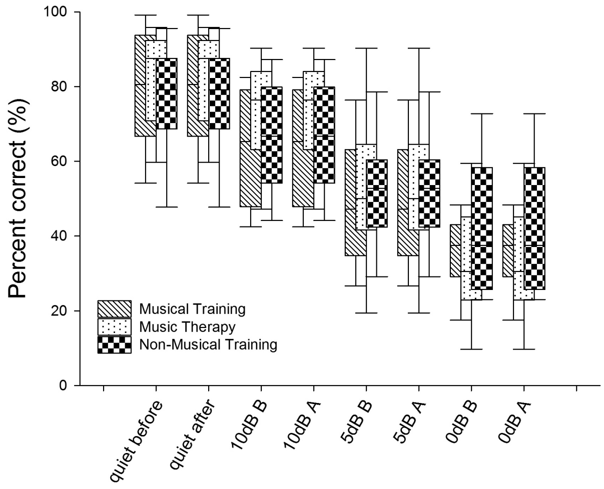


Figure 1. The word identification scores shown for the three different groups in percent correct scores. On the x-axis the letter ‘B’ refers to the results before training; the letter ‘A’ to the results after training.

Sentence identification

Figure 2 shows the results for the sentence identification test before and after the training for the three groups for the three different noise conditions. The figure shows that there are no notable differences between before and after the training, or between the different training groups. In the ANOVA the quiet condition was not used, as it is a percent correct score and not a speech reception threshold. The repeated measures ANOVA was built using the three different SRT scores (stationary, fluctuating and babble) for the three different training groups (musical training, music therapy, non-musical training). The results showed no significant effect within the training groups before and after the training. Also, no differences between the training groups before and after the training were shown.

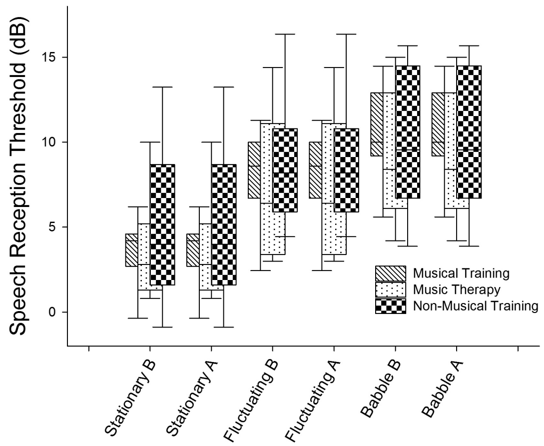


Figure 2. The mean sentence identification is shown in SRT's (dB) plotted for the three training groups. From left to right results are shown for the three different noises: stationary, fluctuating, babble noise, respectively. On the x-axis the letter 'B' refers to the results before training; the letter 'A' to the results after training.

Emotion identification

Figure 3 shows the results for the vocal emotion identification scores before and after the training sessions for the three groups in percent correct scores. The figure shows a possible positive effect of music therapy on the vocal emotion identification. A single, repeated measures ANOVA with the music therapy group and the emotion scores as factors, shows a significant within subject effect ($F(1)=8.898$; $p=0.025$). The repeated measures ANOVA for the three groups did not show an effect between the different training groups. Altogether this could possibly indicate that music therapy could have a possible positive effect on emotion recognition, even though no significant difference between the three groups was shown.

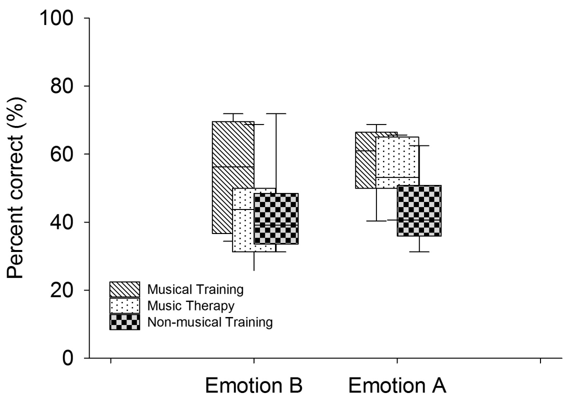


Figure 3. The mean emotion identification shown for the three training groups on the left before and on the right after the training sessions. On the x-axis the letter 'B' refers to the results before training; the letter 'A' to the results after training.

Piano

Figure 4 shows that the musical training group might improve for MCI for piano on the A3 and A5 condition. The other training groups showed no notable improvement after the training for MCI for piano in the figure. The repeated measures ANOVA for the MCI using three piano conditions (no masker, A3, A5) and the three training groups (musical training, music therapy, non-musical training) as factors showed no effect within groups. The between groups ANOVA showed a significant effect of the different training groups ($F(2)=4.481$, $p=0.03$) for MCI. Post-hoc tests showed a significant positive effect of the musical training opposed to the non-musical training ($p=0.03$), but not to the music therapy group. This could indicate that musical training could have a positive effect on melodic contour identification and a better effect than non-musical training, but not better than music therapy.

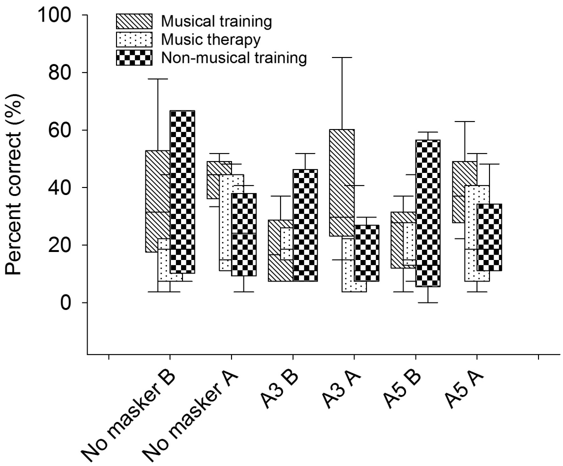


Figure 4. The mean melodic contour identification shown for piano for the three training groups. The three conditions shown from left to right are no masker, A3 masker, and A5 masker. On the x-axis the letter 'B' refers to the results before training; the letter 'A' to the results after training.

Organ

Figure 5 shows that that the musical training group might improve for MCI for organ on the no masker and the A5 condition. The other groups show no notable improvement between before and after the training for MCI for organ in the figure. The repeated measures ANOVA using all three conditions (no masker, A3, A5) and all three groups as factors showed an effect for both MCI ($F(1.97)=6.865$; $p=0.003$) and for the different training groups ($F(3.95)=3.236$; $p=0.025$) within subject. Between subjects an effect for the different groups was shown ($F(2)=5.923$; $p=0.012$). Post-hoc tests indicate a possible effect of musical training opposed to both music therapy ($p=0.038$) and non-musical training ($p=0.018$). This could indicate that musical training could have a positive effect on melodic contour identification for both piano and organ.

9

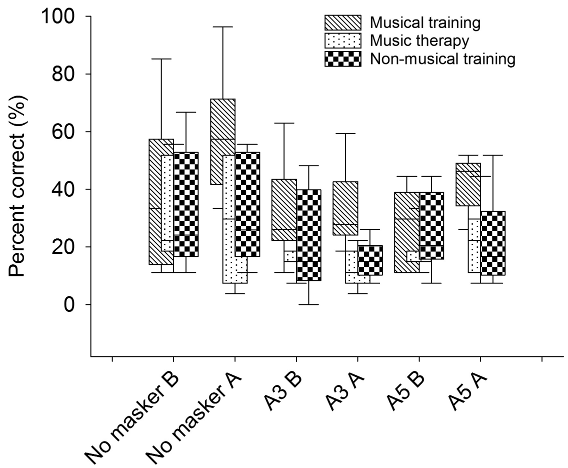


Figure 5. The mean melodic contour identification shown for organ for the three training groups. The three conditions shown from left to right are no masker, A3 masker, and A5 masker. On the x-axis the letter 'B' refers to the results before training; the letter 'A' to the results after training.

Quality of life

Figure 6 shows the total scores before and after the training sessions for the NCIQ. Repeated measures ANOVA showed no effect of the training on the quality of life within and between the three training groups.

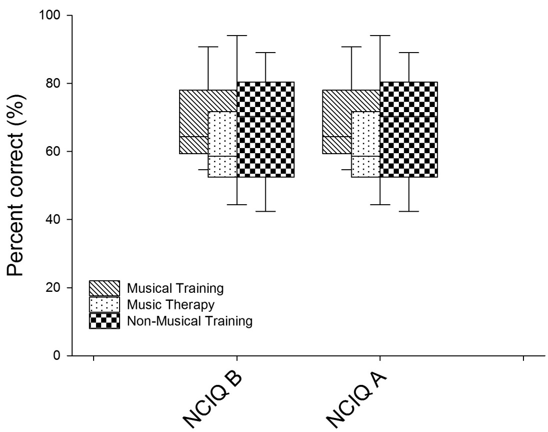


Figure 6. The mean quality of life shown for the three training groups. On the x-axis the letter 'B' refers to the results before training; the letter 'A' to the results after training.

Subjective perceptual skills in the music therapy group

Figure 7 shows the result of the survey in the music therapy group. The figure shows a trend that all participants find themselves improving on all tasks.

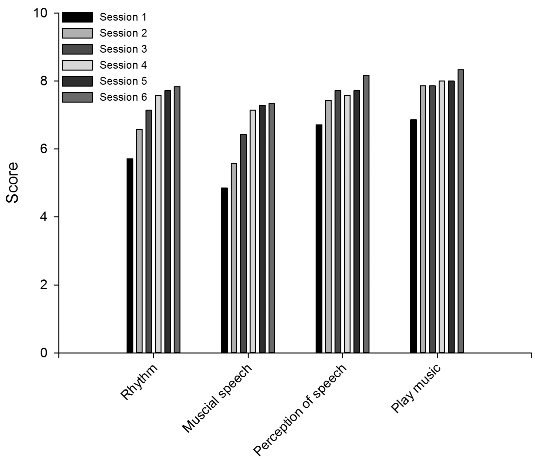


Figure 7. The subjective ability to perceive rhythm, musical speech, music and to play music in the music therapy group per session. The first session on the left, the sixth on the right.

DISCUSSION

The main goal of the current study was to explore the feasibility of implementing a musical training program for CI users in the clinic. A second goal was to investigate differences in training outcomes among different training approaches and to determine which approach was most effective during a short training period. We investigated the effect of three different training approaches (musical training, music therapy and non-musical training) on CI users' auditory perception performance (within and cross domain) and QoL measures. To compare the outcomes of this study to the results for NH musicians listening to CI simulations (Fuller et al. 2014a), the same behavioral tests were used: speech identification, vocal emotion identification and melodic contour identification. A significant within-domain effect (improved melodic contour identification in the musical training group) and a small cross-domain training effect (improved vocal emotion identification in the music therapy group) were observed, but no effects on speech identification (in all three groups) were shown. The music therapy group did show enhanced subjective perceptual skills. No effects on QoL were reported for any of the training groups. A general point of discussion is the small group size that might influence the power and there with the outcome of the current study. We would like to emphasize that the current study was a feasibility study and that future studies should focus on increasing the power by adding more participants, as well as training for longer periods of the time.

Effect of music training on speech perception in noise

No transfer of learning for the three different training groups was found to speech perception (words or sentences). Within this training study, we were not able to replicate the preliminary findings of Patel (2014) and Lo et al. (2015). Patel (2014) showed a small cross-domain benefit for musical training in two CI users from music to speech. One subject showed improved sentence recognition in noise, the other subject improved perception of prosody in speech. Lo et al. (2015) showed a positive effect of musical training on question/statement identification and consonant discrimination in 16 CI users. The lack of strong cross-domain learning in the present study may have been due to the speech outcome measures used. Sentence recognition in noise may depend on perception of voice pitch to some extent, but also involves other high-level cognitive and linguistic processing. The cross-domain music training benefits observed in previous studies may have been due to minimal linguistic context (e.g., consonant identification, as in Lo et al., 2015, syllable perception, as in Zuk, 2013, prosody perception, as in Patel et al., 2014). Perception tests such as emotion identification that explicitly depend on voice pitch perception may have revealed stronger cross-domain training effects (Fuller et al., 2014).

It should be noted that cross-domain music training effects are often small and inconsistent in previous studies. Only a small effect of musical training for speech understanding in noise has been observed in children, young and older adults (Parbery-Clark et al. 2009; Strait and Kraus 2011; Parbery-Clark et al. 2011). Zendel and Alain (2012) showed a significant musician effect for speech perception in noise only for older adults, but no effect for adults younger than 40 years old. Fuller et al. (2014b) showed no effect using normal acoustical stimuli with steady, fluctuating and babble noise, but a small effect using CI simulations. Ruggles et al. (2014) did not find a musician effect for voiced or whispered speech in continuous or gated noise. Larger musician effects have been observed for speech understanding with speech maskers (speech-on-speech). Speech-on-speech perception is an interesting entity to study for potential music training effects. Whereas speech understanding in noise largely involves energetic masking (Gaudrain and Carlyon 2013), speech understanding with competing speech involves both energetic and informational masking, due to lexical content and acoustic similarities between the target and masker talkers (Darwin, Brungart, and Simpson 2003). Voice pitch differences are an important cue for segregating competing talkers. Theoretically, improved pitch perception via musical training would aid talker segregation (Herholz and Zatorre 2012; Kraus, Zatorre, and Strait 2014; Zatorre 2013). Some previous studies have shown significant musician effects for speech-on-speech perception (Swaminathan et al. 2015; Başkent and Gaudrain 2016), while others have not (Boebinger et al. 2015). It is possible that musician effects for speech-on-speech perception may be related to better segregation of other acoustic cues besides voice pitch (Başkent and Gaudrain 2016).

Investigating the musician effect on speech perception in CI users or in NH subjects listening to CI simulations (i.e., under conditions of spectro-temporal degradation) is a relatively understudied and new topic. Our recent research with NH subjects listening to CI simulations showed only a small musician effect for speech perception, significant only for word recognition in noise (Fuller et al. 2014b). Surprisingly, no musician effect on speech perception in noise was observed in NH subjects listening to unprocessed speech. Research thus far has shown mixed results for cross-domain transfer of music training to speech perception in noise. Future research should carefully consider speech outcome measures that depend strongly on perception that is enhanced by music training, most notably pitch perception. If pitch is not a strong cue for a particular speech perception task, then it seems unlikely that music training might benefit that particular speech task. Alternatively, music training may improve working memory and overall pattern perception. Perceptual tasks that explicitly test working memory may further reveal musician advantages.

Effect of music training on emotion identification

As noted above, music training may especially benefit speech perception that depends strongly on voice pitch perception. One such listening task is vocal emotion recognition, which depends strongly on voice pitch perception and is difficult for CI users due to the coarse spectral resolution that does not support harmonic pitch perception (see Moore and Carlyon, 2005 for review). Vocal emotion identification has been shown to be much better in NH listeners than in CI users (House 1994; Pereira 2000; Xin, Fu, and Galvin 2007). Even when listening to acoustic CI simulations with only 4-8 spectral channels, NH listeners have been shown to outperform CI users (Xin, Fu, and Galvin 2007). Gilbers et al. (2015) suggested that NH listeners use mainly the mean pitch for emotion identification, whether listening to unprocessed stimuli or to CI simulations; while real CI users, on the other hand, seemed to rely on pitch ranges conveyed by the temporal modulations. NH musicians have also been shown to outperform NH non-musicians for emotion identification (Thompson, Schellenberg, and Husain 2004).

Our recent CI simulation study in NH musicians and non-musicians showed a significant musician advantage for emotion identification using normal and CI simulated stimuli (Fuller et al. 2014b). Thus, even after degrading the fine structure of the signal musicians were able to identify emotions better. It was suggested that the musician advantage was based on a better perception of pitch cues, even in CI simulations. While pitch perception strongly contributes to emotion identification, other cues that co-vary with F0 also contribute, such as duration (longer for sad, shorter for happy), overall amplitude (higher for happy, lower for sad), tempo and pausing (Luo et al. 2009; Hubbard and Assmann 2013). These co-varying cues were not controlled for in Fuller et al. (2014b) or in the present study. Hence, while emotion identification heavily relies on pitch cues, these other cues may have also contributed to the present pattern of results.

If so, musical training may similarly benefit CI users' vocal emotion identification. However, it should be noted that the extensive training experienced by NH musicians might not be comparable to short-term training with degraded signals in CI users.

In this study, emotion identification was tested using a nonce word, eliminating semantic context cues. We found a positive effect of music therapy on emotion identification, possibly due to improved pitch perception after the training or to improved perception of other co-varying cues such as amplitude and duration. However, no training effect was shown for the other two training groups. One explanation might be that, in contrast to the musical training and the control group, the music therapy group was specifically trained for emotion identification. During the therapy sessions, vocal and instrumental emotion identification was practiced. For example, a member of the group would choose an emotion from a series and play it on an instrument. The other group members' task was to identify the emotion. In the vocal prosody exercise, emotion identification was practiced using a song or a story line that was sung or spoken by the session leaders. These training approaches might have had a direct positive effect on emotion identification in the music therapy group. Our results suggest that emotion identification can be enhanced by direct training with music or speech, and that the interactive nature of the music therapy may have contributed to a better learning of this task. Further research may shed light on the best approach to train CI users' perception of music and pitch-mediated speech.

Effect of training on MCI

It was not surprising that the music training with MCI improved MCI performance. Training benefits were observed for the piano and, to a larger extent, for the organ. Galvin et al. (2008) reported that mean MCI performance in CI users was poorest with piano and best with organ. Other MCI training studies with CI users also showed that performance with the organ improved most with the training (Galvin et al., 2007, 2009). Perhaps, the organ is more easily trained in CI users because its spectral-temporal content is less complex than for other instruments such as the piano. Lo et al (2015) showed that the biggest effect of training occurred after 1 and 2 weeks of training, maximum improvement, however, was seen after 4 to 6 weeks. As we have conducted no intermediate tests during our series, it is unclear whether the training benefit was maximum. Future studies should adopt in-between tests, as well as an extension of the duration of the study to see whether the maximum effect can be found.

No effect on MCI was shown for the music therapy and the control group. Perhaps an extensive training for a specific task creates a big enough benefit in the short period of six weeks, compared to a non-specific training, such as music therapy. Nevertheless an elongation of the training period might show an improvement in the therapy group as well; the elongation might make the total amount of melodic training comparable between groups. Given the generally heterogeneous composition of groups of CI users, future research could focus on larger groups of CI users to be able to draw more definitive conclusions.

Effect of music training on QoL and subjective perception

There was no effect of training on QoL ratings. It is possible that only such a short period of training was not sufficient to alter the QoL. Music therapy has been previously shown to increase subjective QoL ratings in different patient groups (Hilliard 2003; Walworth et al. 2008). But in our music therapy group, no such effect was observed. Anecdotal reports suggested that the music therapy made CI users feel better about their perceptual skills, that they could better understand other talkers' emotions, and that they began to listen to and better enjoy music. This is in line with a recent study by Hütter et al (2015) that showed an increase in subjective overall music perception after therapy. These self-reports of improved speech and music perception are encouraging and should be more deeply investigated in future research.

This feasibility, training study showed an improvement for MCI and emotion identification, only in the groups that were specifically trained for that task. Music therapy positively influenced the subjective perceptual skills of CI users. Our results might indicate that music training or music therapy might be a useful addition to the rehabilitation program of CI users.

Acknowledgement

We would like to thank all CI users for their enthusiasm and commitment with the research and the training sessions. Also thanks to Joeri, Karin, Esmee, Roy, Carmen, Jooske, Han, Angelique, Saar and Aline for the help with the sessions, and to Qian-Jie Fu and the Emily Shannon Foundation for the help with the software and John Galvin for the feedback.

REFERENCES

- Başkent, D., E. Gaudrain, T.N. Tamati, and A. Wagner. 2016. Perception and psychoacoustics of speech in cochlear implant users. . In *Scientific foundations of audiology.*, eds. A. T. Cacace, E. de Kleine, A. Holt and P. van Dijk.
- Başkent, D., and E. Gaudrain. 2016. Musician advantage for speech-on-speech perception. *JASA-EL*.
- Benard, M. R., and D. Başkent. 2013. Perceptual learning of interrupted speech. *PloS One* 8 (3): e58149.
- Besson, M., D. Schon, S. Moreno, A. Santos, and C. Magne. 2007. Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience* 25 (3-4): 399-410.
- Boebinger, D., S. Evans, S.t Rosen, C. F. Lima, T. Manly, and S. K. Scott. 2015. Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America* 137 (1): 378-87.
- Bosman, A. J., and G.F. Smoorenburg. 1995. Intelligibility of dutch CVC syllables and sentences for listeners with normal hearing and with three types of hearing impairment. *Audiology* 34 (5): 260-84.
- Darwin, C. J., D. S. Brungart, and B. D. Simpson. 2003. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America* 114 (5): 2913-22.
- Drennan, W. R., and J. T. Rubinstein. 2008. Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of Rehabilitation Research and Development* 45 (5): 779-89.
- Fu, Q. J, S.l Chinchilla, and J. J. Galvin. 2004. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology* 5 (3): 253-60.
- Fu, Q. J, G. Nogaki, and J. J. Galvin III. 2005. Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation. *Journal of the Association for Research in Otolaryngology* 6 (2): 180-9.
- Fu, Q. J., and J. J. Galvin 3rd. 2008. Maximizing cochlear implant patients' performance with advanced speech training procedures. *Hearing Research* 242 (1-2) (Aug): 198-208.
- Fuller, C. D., R. H. Free, A. Maat, and D. Başkent. under revision. Music and quality of life in post-lingually deafened cochlear implant users. Fuller, C. D., J. J. Galvin 3rd, R. H. Free, and D. Başkent. 2014a. Musician effect in cochlear implant simulated gender categorization. ***The Journal of the Acoustical Society of America*** 135 (3) (Mar): EL159-65.
- Fuller, C. D., J. J. Galvin 3rd, B. Maat, R. H. Free, and D. Başkent. 2014b. The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience* 8 (Jun 30): 179.
- Fuller, C. D., E. Gaudrain, J. N. Clarke, J. J. Galvin, Q. J. Fu, R. H. Free, and D. Başkent. 2014c. Gender categorization is abnormal in cochlear implant users. *Journal of the Association for Research in Otolaryngology* 15 (6) (Dec): 1037-48.
- Galvin, J. J., E. Eskridge, S. Oba, and Q. J Fu. 2012. Melodic contour identification training in cochlear implant users with and without a competing instrument. Paper presented at Seminars in hearing, .
- Galvin, J. J.,3rd, Q. J. Fu, and G. Nogaki. 2007. Melodic contour identification by cochlear implant listeners. *Ear and Hearing* 28 (3) (Jun): 302-19.
- Galvin, J. J.,3rd, Q. J. Fu, and S. Oba. 2008. Effect of instrument timbre on melodic contour identification by cochlear implant users. *The Journal of the Acoustical Society of America* 124 (4) (Oct): EL189-95.
- Galvin, J. J.,3rd, Q. J. Fu, and S. I. Oba. 2009. Effect of a competing instrument on melodic contour identification by cochlear implant users. *The Journal of the Acoustical Society of America* 125 (3) (Mar): EL98-103.
- Gaudrain, E., and R. P. Carlyon. 2013. Using zebra-speech to study sequential and simultaneous speech segregation in a cochlear-implant simulation. *The Journal of the Acoustical Society of America* 133 (1): 502-18.
- Gfeller, K., A. Christ, J. F. Knutson, S. Witt, K. T. Murray, and R. S. Tyler. 2000. Musical backgrounds, listening

habits, and aesthetic enjoyment of adult cochlear implant recipients. *Journal of the American Academy of Audiology* 11 (7) (Jul-Aug): 390-406.

- Gfeller, K., C. Turner, M. Mehr, G. Woodworth, R. Fearn, J. F. Knutson, S. Witt, and J. Stordahl. 2002a. Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants International* 3 (1) (Mar): 29-53.
- Gfeller, K., C. Turner, J. Oleson, X. Zhang, B. Gantz, R. Froman, and C. Olszewski. 2007. Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear and Hearing* 28 (3) (Jun): 412-23.
- Gfeller, K., S. Witt, M. Adamek, M. Mehr, J. Rogers, J. Stordahl, and S. Ringgenberg. 2002b. Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *Journal of the American Academy of Audiology* 13 (3) (Mar): 132-45.
- Gilbers, S., C. Fuller, D. Gilbers, M. Broersma, M. Goudbeek, R. Free, and D. Başkent. 2015. Normal-hearing listeners' and cochlear implant users' perception of pitch cues in emotional speech. *I-Perception* 6 (5): 0301006615599139.
- Goudbeek, M., and M. Broersma. 2010. The demo/kemo corpus: A principled approach to the study of cross-cultural differences in the vocal expression and perception of emotion. Paper presented at 7th International Conference on Language Resources and Evaluation (LREC 2010), .
- Herholz, S. C., and R. J. Zatorre. 2012. Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron* 76 (3) (Nov 8): 486-502.
- Hilliard, R. E. 2003. The effects of music therapy on the quality and length of life of people diagnosed with terminal cancer. *Journal of Music Therapy* 40 (2) (Summer): 113-37.
- Hinderink, J. B., P. F. Krabbe, and P. Van Den Broek. 2000. Development and application of a health-related quality-of-life instrument for adults with cochlear implants: The nijmegen cochlear implant questionnaire. *Otolaryngology--Head and Neck Surgery* 123 (6) (Dec): 756-65.
- Hubbard, D. J., and P. F. Assmann. 2013. Perceptual adaptation to gender and expressive properties in speech: The role of fundamental frequency. *The Journal of the Acoustical Society of America* 133 (4) (Apr): 2367-76.
- Hütter, E., H. Argstatter, M. Grapp, and P. K. Plinkert. 2015. Music therapy as specific and complementary training for adults after cochlear implantation: A pilot study. *Cochlear Implants International* 16 Suppl 3 (Sep): S13-21.
- Ingvalson, E. M., B. Lee, P. Fiebig, and P. C. Wong. 2013. The effects of short-term computerized speech-in-noise training on postlingually deafened adult cochlear implant recipients. *Journal of Speech, Language, and Hearing Research : JSLHR* 56 (1) (Feb): 81-8.
- Kong, Y. Y., R. Cruz, J. A. Jones, and F. G. Zeng. 2004. Music perception with temporal cues in acoustic and electric hearing. *Ear and Hearing* 25 (2) (Apr): 173-85.
- Kong, Y. Y., J. M. Deeks, P. R. Axon, and R. P. Carlyon. 2009. Limits of temporal pitch in cochlear implants. *The Journal of the Acoustical Society of America* 125 (3) (Mar): 1649-57.
- Kraus, N., R. J. Zatorre, and D. L. Strait. 2014. Editors' introduction to hearing research special issue: Music: A window into the hearing brain. *Hearing Research* 308 (Feb): 1.
- Limb, C. J., and A. T. Roy. 2014. Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing Research* 308 (Feb): 13-26.
- Lo, C. Y., C. M. McMahon, V. Looi, and W. F. Thompson. 2015. Melodic contour training and its effect on speech in noise, consonant discrimination, and prosody perception for cochlear implant recipients. *Behavioural Neurology* 2015 : 352869.
- Loebach, J. L., and D. B. Pisoni. 2008. Perceptual learning of spectrally degraded speech and environmental sounds. *The Journal of the Acoustical Society of America* 123 (2) (Feb): 1126-39.
- Looi, V., and J. She. 2010. Music perception of cochlear implant users: A questionnaire, and its implications for a music training program. *International Journal of Audiology* 49 (2): 116-28.

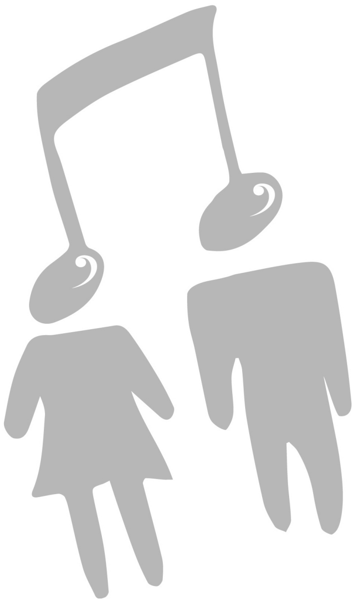
- Looi, V., K. Gfeller, and V. Driscoll. 2012. Music appreciation and training for cochlear implant recipients: A review. *Seminars in Hearing* 33 (4) (Nov 1): 307-34.
- Luo, X., Q. J Fu, H. P. Wu, and C. J. Hsu. 2009. Concurrent-vowel and tone recognition by mandarin-speaking cochlear implant users. *Hearing Research* 256 (1): 75-84.
- Marques, C., S. Moreno, S. L. Castro, and M. Besson. 2007. Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience* 19 (9): 1453-63.
- McDermott, H. J. 2004. Music perception with cochlear implants: A review. *Trends in Amplification* 8 (2): 49-82.
- Migchelbrink, F.G.H.M., and F.Brinkman. 2000. *Praktijkgericht onderzoek in zorg en welzijn*SWP.
- Nogaki, G., Q. J. Fu, and J. J. Galvin 3rd. 2007. Effect of training rate on recognition of spectrally shifted speech. *Ear and Hearing* 28 (2) (Apr): 132-40.
- Oba, S. I., Q. J. Fu, and J. J. Galvin 3rd. 2011. Digit training in noise can improve cochlear implant users' speech understanding in noise. *Ear and Hearing* 32 (5) (Sep-Oct): 573-81.
- Pacchetti, C., F. Mancini, R. Aglieri, C. Fundaro, E. Martignoni, and G. Nappi. 2000. Active music therapy in parkinson's disease: An integrative method for motor and emotional rehabilitation. *Psychosomatic Medicine* 62 (3) (May-Jun): 386-93.
- Parbery-Clark, A., D. L. Strait, S. Anderson, E. Hittner, and N. Kraus. 2011. Musical experience and the aging auditory system: Implications for cognitive abilities and hearing speech in noise. *PLoS One* 6 (5): e18082.
- Parbery-Clark, A., E. Skoe, C. Lam, and N. Kraus. 2009. Musician enhancement for speech-in-noise. *Ear and Hearing* 30 (6) (Dec): 653-61.
- Patel, A. D. 2014. Can nonlinguistic musical training change the way the brain processes speech? the expanded OPERA hypothesis. *Hearing Research* 308 (Feb): 98-108.
- Peterson, G. E., and H. L. Barney. 1952. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America* 24 (2): 175-84.
- Philips, B., B. Vinck, E. De Vel, L. Maes, W. D'Haenens, H. Keppler, and I. Dhooge. 2012. Characteristics and determinants of music appreciation in adult CI users. *European Archives of Oto-Rhino-Laryngology* 269 (3) (Mar): 813-21.
- Plomp, R., and A. M. Mimpen. 1979. Speech-reception threshold for sentences as a function of age and noise level. *The Journal of the Acoustical Society of America* 66 (5): 1333-42.
- Ruggles, D. R., R. L. Freyman, and A. J. Oxenham. 2014. Influence of musical training on understanding voiced and whispered speech in noise. *PloS One* 9 (1) (Jan 28): e86980.
- Smith, D. R. R., and R. D. Patterson. 2005. The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and agea). *The Journal of the Acoustical Society of America* 118 (5): 3177-86.
- Stacey, P. C., and A. Q. Summerfield. 2007. Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. *The Journal of the Acoustical Society of America* 121 (5 Pt1) (May): 2923-35.
- Stacey, P. C., C. H. Raine, G. M. O'Donoghue, L. Tapper, T. Twomey, and A. Q. Summerfield. 2010. Effectiveness of computer-based auditory training for adult users of cochlear implants. *International Journal of Audiology* 49 (5) (May): 347-56.
- Stacey, P. C., and A. Q. Summerfield. 2008. Comparison of word-, sentence-, and phoneme-based training strategies in improving the perception of spectrally distorted speech. *Journal of Speech, Language, and Hearing Research : JSLHR* 51 (2) (Apr): 526-38.
- Strait, D. L., and N. Kraus. 2011. Can you hear me now? musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *The Relationship between Music and Language*: 99.
- Swaminathan, J., C. R. Mason, T. M. Streeter, V. Best, G. Kidd Jr, and A. D. Patel. 2015. Musical training,

individual differences and the cocktail party problem. *Scientific Reports* 5 .

- Van de Winckel, A., H. Feys, W. De Weerdt, and R. Dom. 2004. Cognitive and behavioural effects of music-based exercises in patients with dementia. *Clinical Rehabilitation* 18 (3) (May): 253-60.
- Walworth, D., C. S. Rumana, J. Nguyen, and J. Jarred. 2008. Effects of live music therapy sessions on quality of life indicators, medications administered and hospital length of stay for patients undergoing elective surgical procedures for brain. *Journal of Music Therapy* 45 (3) (Fall): 349-59.
- Wright, R., and R. M. Uchanski. 2012. Music perception and appraisal: Cochlear implant users and simulated cochlear implant listening. *Journal of the American Academy of Audiology* 23 (5) (May): 350,65; quiz 379.
- Xin, L., Q. J. Fu, and J. J. Galvin 3rd. 2007. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification* 11 (4) (Dec): 301-15.
- Yang, L. P., and Q. J. Fu. 2005. Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *The Journal of the Acoustical Society of America* 117 (3): 1001-4.
- Zatorre, R. J. 2013. Predispositions and plasticity in music and speech learning: Neural correlates and implications. *Science (New York, N.Y.)* 342 (6158) (Nov 1): 585-9.
- Zendel, B. R., and C. Alain. 2013. The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *Journal of Cognitive Neuroscience* 25 (4): 503-16.
- Zendel, B. R., and C. Alain. 2012. Musicians experience less age-related decline in central auditory processing. *Psychology and Aging* 27 (2): 410.

Chapter 10

General discussion and conclusions



This thesis focused on perception of music and speech in CI users. While (re-)gaining speech perception is the most important goal for CI users, music perception is also highly valued (Drennan and Rubinstein 2008). However, music perception and enjoyment is often poor in CI users (Limb and Roy 2014; McDermott 2004). This thesis first described music perception in terms of subjective and behavioral measures in a large cohort of CI users. The subjective measures were important because music ultimately conveys emotion, and the effect of music on one's emotional state may differ from one individual to another (Zatorre and Salimpoor 2013; Salimpoor et al. 2009). The behavioral measures were important to quantify CI users' music perception, which in turn is important to guide development of future devices and signal processing strategies. Theoretically, improvements in music perception may also improve music enjoyment. Improved pitch perception would most likely also improve perception of indexical cues (who is talking) and prosodic cues (how something is said) in speech, as subjective perception of music and pitch-mediated speech were correlated in our research. While it may be difficult to sufficiently improve CI technology to support good pitch perception, we secondly explored methods in the training domain that might improve perception of music and pitch-mediated speech tasks. A strong support for this idea came from the studies with both musicians who were trained previously and also with CI users who were trained directly.

One of the first parts of the thesis was an exploration of the subjective appreciation and perception of music within CI users. We hypothesized that music perception and/or enjoyment would be related to speech perception and/or quality of life (QoL). We not only had access to typical CI users (post-lingually deafened), but also to an under-studied CI group: early-deafened, late-implanted (EDLI). The comparison between these two groups revealed many important factors that can affect music appreciation and perception with CIs.

Post-lingually deafened CI users (the dominant population of CI patients) reported that the sound quality of music was poor **(Chapter 3)**. Perception of music has been reported to be poorer and less enjoyable in post-lingually deafened CI-users (Gfeller et al. 2000; Lassaletta et al. 2008; McDermott 2004; Limb and Roy 2014). Our findings re-emphasize the need of improvement of the perception of music for post-lingually deafened CI users. In contrast, EDLI CI users reported that the sound quality of music was good **(Chapter 2**), though still suboptimal. This observation in EDLI CI users is in agreement with Trehub, Vongpaisal, and Nakata (2009), who reported high music appreciation levels in pediatric CI users even though their musical pitch perception was poorer than their NH peers. These results suggest that music perception may not be the only (or even main) determinant of music appreciation, and that the amount of acoustic hearing experience before implantation may greatly affect music sound quality perception.

The relationship between subjective perception of music and quality of life (QoL) also differed between these two groups of CI users. In post-lingually deafened CI users, music enjoyment and perception was related to QoL **(Chapter 3)**, similar to findings by Lassaletta et al. (2007) who showed that music appreciation was correlated to the generic QoL in a smaller number of post-lingually deafened CI-users. However, we found no significant correlation between music appreciation and QoL in EDLI CI users **(Chapter 2)**. This difference may be because of differences in music appreciation between post-lingually deafened and EDLI CI users, with "perfect" perception of music being more important to post-lingually deafened CI users, based on their previous acoustic experience with music. Providing rehabilitation to post-lingually deafened CI users might increase their QoL. There was no association between behaviorally measured speech perception and subjective music perception for either group. A significant effect of the perception of the elements of music and subjective measure of hearing abilities (the SSQ questionnaire) was observed only in the post-lingually deafened CI users. Philips et al. (2012) did show a correlation between a more natural sounding music and speech reception thresholds in quiet and noise, in a small group of post-lingually deafened CI users. Even though the evidence from the subjective studies is weak, improving the perception and enjoyment of music could possibly affect both QoL and speech perception. Further research needs to be conducted to determine whether better music perception is linked to higher QoL and better speech perception in CI users.

In terms of rating the subjective perception of the music elements, post-lingually deafened CI users rated rhythm as the easiest element and melody as most difficult element of music to perceive, in accordance with previous literature (Gfeller et al. (2000); **Chapter 3**). In contrast, EDLI CI users rated melody the easiest and rhythm the most difficult element **(Chapter 2)**. One explanation for these contradictory patterns of results is that, different from post-lingually deafened CI users, EDLI CI users have no previous acoustic listening experience with music. EDLI CI users may develop music concepts and patterns differently with electric hearing, which might differ from the normal, acoustic listening experience. Given the limits of pitch perception with the CI, it is curious that EDLI CI users would rate melody higher than rhythm elements. While pitch is represented different and more coarsely in electric hearing than in acoustic hearing, EDLI CI users seemed able to develop meaningful melodic patterns; given their previous acoustic hearing, post-lingually deafened CI users may have greater difficulty adapting to melodic patterns with electric hearing. Previous studies have also shown that amplitude modulation detection is poorer in pre-lingually than in post-lingually deafened CI users (De Ruiter et al. 2015). If temporal envelope processing was poorer, EDLI CI users may have had greater difficulty with rhythm perception than did the post-lingually deafened CI users.

The subjective evaluation of music showed that both perception and enjoyment remained unsatisfactory in both CI subject groups. Because music perception (and thus, appreciation) is limited by the poor pitch perception provided by CIs, we investigated how this limitation might also affect perception of pitch mediated speech. Our research showed that post-lingually deafened CI users exhibited abnormal voice gender categorization **(Chapter 4)** and vocal emotion identification **(Chapter 5)**, relative to NH listeners. Our findings extended those from previous literature (Xin, Fu, and Galvin 2007; Winn, Chatterjee, and Idsardi 2011), and revealed further details regarding CI users' perception of voice gender and vocal emotion, suggesting that the extent of the problem may be larger than what was previously reported.

In NH listeners, voice gender categorization is mostly based on perception of F0 and vocal tract length (VTL), although there are other acoustic cues such as breathiness (Holmberg, Hillman, and Perkell 1988; Van Borsel, Janssens, and De Bodt 2009) or intonation (Fitzsimons, Sheahan, and Staunton 2001). In our study, NH listeners indeed made effective use of both VTL and F0 cues, even when listening to the spectro-temporally degraded CI simulations. Our findings showed, for the first time, that CI users almost exclusively relied on F0 cues, and did not make use of the VTL cues for voice gender categorization. This finding seems at first surprising as there is a long history of research that focused on deficiencies of voice pitch (F0) perception in CI users (Kong and Carlyon 2010; Oxenham 2008; Kong et al. 2009; Gfeller et al. 2007; Fu, Chinchilla, and Galvin 2004; Başkent et al. 2016). Only after the systematic manipulation of the voices, as we did in this study, it was revealed that CI users can utilize F0 cues, even if they are only weakly delivered by the device. What seems to be more problematic is the perception of VTL cues, which has been rarely studied in CI listeners.

Given that the spectro-temporal degradation was present in the signal of both the real CIs and the CI simulations, the difference in VTL perception between NH and CI listeners may be due to properties of electric stimulation. A study by Gaudrain and Başkent (2015) showed no differences for the just-noticeable-difference between VTL and F0 perception using a fixed number of sinewave vocoded channels in NH listeners. Changing the number of channels had a bigger effect on VTL than on F0 perception. Given the limited spectral resolution (due to current spread among the small number of implanted electrodes), an improvement in VTL perception may not be feasible in CI users (Gaudrain and Başkent 2015). Another explanation might be that CI users do perceive a weak VTL, but are unable to use it. If the latter is true a musical training providing for example timbre or VTL cues, might enable usage of VTL and therewith improve gender recognition. Gender categorization is found to be abnormal in CI users in comparison to NH listeners, due to different weighting of voice cues. Future studies could focus on adding other acoustic cues (e.g., breathiness or intonation) to the analyses to further unravel differences in perception between CI users and NH listeners.

Chapter 5 focused on another pitch-based speech task: vocal emotion identification. Emotions can be divided according to arousal and valance. Arousal differentiates emotions with a high (anger) or low arousal (sad). Valence differentiates positive (happy) from negative emotions (sad). Vocal emotion identification is mostly based on arousal in general (Russell and Mehrabian 1977). In our vocal emotion identification study, we used pitch range and mean pitch as measures of arousal based on previous research (Xin, Fu, and Galvin 2007; Goudbeek and Broersma 2010). Results showed that NH listeners used mean pitch for emotion identification for both unprocessed stimuli and acoustic CI simulations. CI users relied on pitch ranges, and did not cue to mean pitch. CI users might not be able to use mean pitch as a cue due to the poor spectral resolution. Instead they appeared to utilize the more easily perceived pitch range cue. CI users had a poorer vocal emotion identification compared to NH subjects listening to unprocessed speech or CI simulations. These findings are in accordance to other studies that showed CI users to have a different identification of emotions based on the cues available via the CI (Winn, Chatterjee, and Idsardi 2011; Xin, Fu, and Galvin 2007).

The impaired subjective and behavioral perception of pitch-related speech and music perception as described in **Chapters 2, 3, 4 and 5** re-emphasized the need to improve pitch perception in CI users. One way to improve music perception (and perhaps pitch perception) might be through musical training. NH musicians have been shown to have advantages over non-musicians in perception of music, as would be expected. More interestingly, musicians also have been shown to perform better in some speech-related tasks, displaying a cross-domain transfer of learning (Parbery-Clark et al. 2009; Schon, Magne, and Besson 2004; Besson et al. 2007; Thompson, Schellenberg, and Husain 2004; Chartrand and Belin 2006; Patel 2014; Micheyl et al. 2006; Kraus and Chandrasekaran 2010; Kraus, Zatorre, and Strait 2014). Previous studies in NH listeners have shown the largest musician effect for speech understanding with speech maskers, with smaller or no effects for speech in noise (Ruggles, Freyman, and Oxenham 2014; Parbery-Clark et al. 2009; Boebinger et al. 2015; Swaminathan et al. 2015; Zendel and Alain 2013). Several hypotheses have been developed about what specifically may be helping speech perception from music training. Better perception of auditory cues may provide a bottom-up advantage (Herholz and Zatorre 2012; Zatorre and Baum 2012; Micheyl et al. 2006). Alternatively, enhanced auditory cognitive functioning provide a top-down advantage (Strait et al. 2010; Kraus, Zatorre, and Strait 2014; Zendel and Alain 2013).

Both advantages provide a benefit when listening to spectro-temporally degraded music and speech, as is the case in CI-processed signals. Ideally, this idea would be tested with musically trained CI users. However, it was difficult to find CI users with extensive musical experience before and after implantation in our medical center **(Chapter 6)**. To explore whether musical experience might contribute to spectro-temporally degraded music and speech perception, we tested NH musicians and provided training directly to actual CI users. NH musicians were used as a model of long-term musical training.

To explore the musician effect on degraded music and speech perception, we have recruited NH musicians (defined as having 10 or more years of musical training) and non-musicians **(Chapters 7 and 8)**. We have tested perception of music (melodic contour identification), speech (sentence recognition in quiet and in noise), and pitch-mediated speech (vocal emotion identification, voice gender categorization). Spectro-temporal degradation was implemented via 8-channel acoustic CI simulations. For music, the degradation significantly reduced the performance in NH musicians and non-musicians; musicians exhibited better music performance than did non-musicians. This confirmed that within-domain learning effects from musical training are robust, and preserved even when heavy degradations are imposed on the music stimuli.

NH musician performance was very good for vocal emotion identification and voice gender categorization. A cross-domain musician effect for pitch-related speech tasks was observed in the CI simulations. **Chapter 5** showed that NH listeners made use of the mean pitch for emotion identification compared to CI users. NH musicians seemed better able to perceive mean pitch in the CI simulations and performed better than non-musicians. **Chapter 7** showed musicians were able to better utilize F0 cues for voice gender categorization. If a better perception of pitch is the underlying mechanism for the musician effect in pitch-related speech tasks, this could imply that musical training with targeted pitch perception could enhance CI users' emotion identification and voice gender categorization. For speech perception, the picture was more complicated. We found no musician advantage for intelligibility with unprocessed speech, and only a small advantage for word identification in one noise condition with the CI simulation. These results are in agreement with previous studies that found small or inconsistent transfer of the musician effect to speech tasks (Ruggles, Freyman, and Oxenham 2014; Parbery-Clark et al. 2009; Boebinger et al. 2015; Swaminathan et al. 2015; Zendel and Alain 2013). The present results add to body of literature showing a musician advantage for pitch-mediated speech perception, even under conditions of spectro-temporal degradation **(Chapter 7 and 8)**. The contribution from other skills potentially improved by music training (e.g., better segregation of stimuli, which was not studied here), remain to be determined.

We note that sinewave vocoders were used for the CI simulations to study the musician effect. Sinewave vocoders were used because they provide better representation of the temporal envelope, compared to noise-band vocoders (e.g., Fu et al. 2005). Alternatively, "low-noise" noiseband vocoders have been shown to improve envelope fidelity in CI simulations (Whitmal et al. 2007). One disadvantage associated with sinewave vocoding is the proliferation of side bands around the carrier frequency that might provide additional spectral cues that might benefit pitch perception (Souza and Rosen 2009). In this study, such sideband cues were available to both the NH musicians and non-musicians; only the NH musicians seemed able to use this cue if it was indeed meaningful. Sinewave carriers also create the perception of a somewhat constant pitch; spectral envelope information (the relative amplitudes across the sinewave carriers) may have been better perceived by the NH musicians. The sinewave vocoders implemented in these studies did not simulate channel interaction that typically exists in CI users. Crew et al. (2012) used a channel mixing technique to introduce different degrees of channel interaction in sinewave vocoded CI simulations, and found that melodic pitch perception worsened as the channel interaction increased. Also, only 8-channel vocoders were used for the CI simulations. CI users' functional spectral resolution has been shown to range from 4-12 channels (Friesen et al. 2001; Xu, Thompson, and Pfingst 2005). It is unclear whether increasing or decreasing the number of channels in the present CI simulations would have enhanced or reduced the musician effect. There are also tradeoffs between the number of channels and then temporal envelope cutoff frequency (160 Hz in this study). When there are fewer channels, listeners can utilize higher frequency temporal envelope information. However, when there are a sufficient number of spectral channels, the temporal envelope filter can be reduced to 50 Hz or less with no effect on performance (Fu et al., 2004). Performance in the tasks used in these studies (vocal emotion identification, gender categorization, melodic contour identification) partly depended on temporal and spectral envelope cues. Future studies may manipulate these cues (number of channels, vocoder carriers, temporal envelope cut-off frequency, etc.) to further explore the musician effect with CI simulations.

In general, the results from this thesis showed that the musician advantage was stronger as the importance of pitch in the listening task increased. This supports the idea that the musician effect is strongly rooted in enhanced pitch perception. If musician advantage was mainly due to better cognitive processing (which was not tested here but in other studies (Zendel and Alain 2013; Kraus, Zatorre, and Strait 2014; Herholz and Zatorre 2012; Zatorre 2013), the musicians in our study would have performed better in all perception tasks, which was not the case. The results suggest that musical training before (and possibly after) implantation might offer some advantage in pitch processing that may persist under conditions of spectro-temporal degradation. This advantage might strongly benefit perception of prosodic and indexical cues in speech, as well as melodic pitch in music.

In the last part of the thesis, we implemented a feasibility study for training actual CI users

(Chapter 9). as a first step toward implementing a musical training program in the clinic. This study explored whether any training improvements could be observed in such a short period of time (6 weeks), and if so, which training approach was most effective. A small group of CI users, representing a typical group of older, post-lingually deafened CI users, was trained for six weeks. Three different training methods were compared: individualized musical training, group music therapy and group non-musical training. To be consistent with our previous study with NH musicians and non-musicians **(Chapter 8)**, CI users were tested on speech intelligibility, vocal emotion recognition, and melodic contour identification before and after training. Similar to NH musicians, a benefit for the trained task was also observed in CI users; post-training improvements in melodic contour identification were observed for the individualized musical training group. What was perhaps more interesting is that post-training improvements were also observed in the music therapy group for vocal emotion identification, a pitch related speech task. This suggests that a specific task can be trained in a short period of time, as both melodic contour and emotion identification were trained in the musical training and the music therapy group, respectively. Similar to NH musicians, perception of pitch-mediated speech can be trained in CI users. Future studies may include longer training periods and intermittent behavioral testing during training.

There was no transfer of training benefit for speech intelligibility for any of the three training methods. These findings are in agreement with the mixed results observed in previous studies (Patel 2014; Lo et al. 2015). However, it is still unclear whether improvements might be observed with a larger subject group or a longer training period. We find it promising that at least one training approach (music therapy) improved one non-music task (vocal emotion identification). Supporting this optimism, the CI users in the music therapy group reported a subjective improvement in music perception, in line with recent literature (Hütter et al. 2015). Together with the musician effect observed in NH listeners, these findings have positive implications for music training in CI users. Music training could improve CI users' music perception (and maybe also speech perception), which in turn could improve music enjoyment (Fuller et al. 2013), which in turn could improve QoL (Fuller et al. under revision).

Overall, our findings with musicians suggest that musical experience before (and possibly after) implantation might improve CI outcomes. This might be especially true for children implanted at a very young age, who develop speech and music patterns exclusively via electric hearing during the most optimal years of neuroplasticity (Houston and Miyamoto 2010; Tajudeen et al. 2010; Olszewski et al. 2005; Magne, Schon, and Besson 2006; Besson et al. 2007). Previous studies with adult CI users have shown that computer-based training can significantly improve music perception (Galvin et al. 2012; Galvin, Fu, and Nogaki 2007). Computer-based music training has also been shown to improve Mandarin-speaking pediatric CI users' music and lexical tone perception (Fu et al. 2015). Music training might be especially effective for bimodal CI listeners (i.e., combined use of CI and hearing aid) who combine fine-structure cues from acoustic hearing with envelope cues from electric hearing (Cullington and Zeng 2011; Sucher and Mcdermott 2009). As the biggest effect was shown for music and pitch-related speech tasks (emotion identification and vocal gender categorization), an effect of training might be largest for these tasks. The effect on speech perception is unclear, as our studies combined with previous literature, show small to no effects form musical training on speech intelligibility.

REFERENCES

- Başkent, D., E. Gaudrain, TN Tamati, and A. Wagner. 2016. Perception and psychoacoustics of speech in cochlear implant users. In *Scientific foundations of audiology.*, eds. A. T. Cacace, E. de Kleine, A. Holt and P. van Dijk.
- Besson, M., D. Schon, S. Moreno, A. Santos, and C. Magne. 2007. Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience* 25 (3-4): 399-410.
- Boebinger, D., Evans, S., Rosen, S., Lima, C.F., Manly, T., and Scott, S.K. 2015. Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America* 137 (1): 378-87.
- Crew JD, Galvin JJ 3rd, Fu QJ. 2012 Channel interaction limits melodic pitch perception in simulated cochlear implants. J Acoust Soc Am. 2012 Nov;132(5):EL429-35
- Chartrand, J. P., and P. Belin. 2006. Superior voice timbre processing in musicians. *Neuroscience Letters* 405 (3) (Sep 25): 164-7.
- Cullington, H. E., and F. G. Zeng. 2011. Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination, and talker identification. *Ear and Hearing* 32 (1) (Feb): 16-30.
- De Ruiter, A. M., J. A. Debruyne, M. N. Chenault, T. Francart, and J. P. Brokx. 2015. Amplitude modulation detection and speech recognition in late-implanted prelingually and postlingually deafened cochlear implant users. *Ear and Hearing* 36 (5) (Sep-Oct): 557-66.
- Drennan, W. R., and J. T. Rubinstein. 2008. Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of Rehabilitation Research and Development* 45 (5): 779-89.
- Fitzsimons, M., N. Sheahan, and H. Staunton. 2001. Gender and the integration of acoustic dimensions of prosody: Implications for clinical studies. *Brain and Language* 78 (1): 94-108.
- Friesen, L. M., R. V. Shannon, D. Başkent, and X. Wang. 2001. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America* 110 (2) (Aug): 1150-63.
- Fu, Q.J., S. Chinchilla, and J. J. Galvin. 2004. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *Journal of the Association for Research in Otolaryngology* 5 (3): 253-60.
- Fu, Q. J, G. Nogaki, and J. J. Galvin III. 2005. Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation. *Journal of the Association for Research in Otolaryngology* 6 (2): 180-9.
- Fu, Q.J., J. J. Galvin, Xiaosong Wang, and Jiunn-Liang Wu. 2015. Benefits of music training in mandarin-speaking pediatric cochlear implant users. *Journal of Speech, Language, and Hearing Research* 58 (1): 163-9.
- Fuller, C., L. Mallinckrodt, B. Maat, D. Başkent, and R. Free. 2013. Music and quality of life in early-deafened late-implanted adult cochlear implant users. *Otology & Neurotology* 34 (6) (Aug): 1041-7.
- Fuller, C. D., R. H. Free, A. Maat, and D. Başkent. under revision. Music and quality of life in post-lingually deafened cochlear implant users.
- Galvin, J. J., E. Eskridge, S. Oba, and Q.J. Fu. 2012. Melodic contour identification training in cochlear implant users with and without a competing instrument. Paper presented at Seminars in hearing, .
- Galvin, J. J.,3rd, Q. J. Fu, and G. Nogaki. 2007. Melodic contour identification by cochlear implant listeners. *Ear and Hearing* 28 (3) (Jun): 302-19.
- Gaudrain, E., and D. Başkent. 2015. Factors limiting vocal-tract length discrimination in cochlear implant simulations. *The Journal of the Acoustical Society of America* 137 (3) (Mar): 1298-308.
- Gfeller, K., A. Christ, J. F. Knutson, S. Witt, K. T. Murray, and R. S. Tyler. 2000. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *Journal of the American Academy of*

Audiology 11 (7) (Jul-Aug): 390-406.

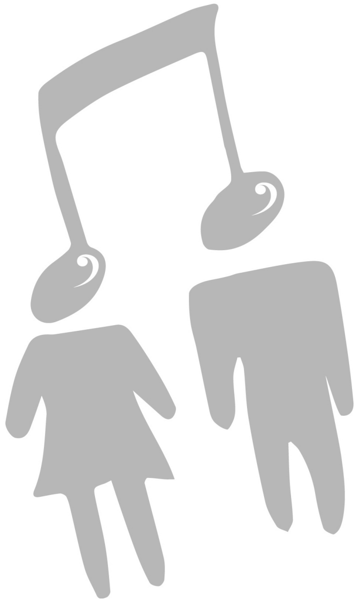
- Gfeller, K., C. Turner, J. Oleson, X. Zhang, B. Gantz, R. Froman, and C. Olszewski. 2007. Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise. *Ear and Hearing* 28 (3) (Jun): 412-23.
- Goudbeek, M., and M. Broersma. 2010. Language specific effects of emotion on phoneme duration. Paper presented at INTERSPEECH, .
- Herholz, S. C., and R. J. Zatorre. 2012. Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron* 76 (3) (Nov 8): 486-502.
- Holmberg, E. B., R. E. Hillman, and J. S. Perkell. 1988. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *The Journal of the Acoustical Society of America* 84 (2): 511-29.
- Houston, D. M., and R. T. Miyamoto. 2010. Effects of early auditory experience on word learning and speech perception in deaf children with cochlear implants: Implications for sensitive periods of language development. *Otology & Neurotology.*31 (8) (Oct): 1248-53.
- Hütter, E., H. Argstatter, M. Grapp, and P. K. Plinkert. 2015. Music therapy as specific and complementary training for adults after cochlear implantation: A pilot study. *Cochlear Implants International* 16 Suppl 3 (Sep): S13-21.
- Kong, Y. Y., and R. P. Carlyon. 2010. Temporal pitch perception at high rates in cochlear implants. *The Journal of the Acoustical Society of America* 127 (5) (May): 3114-23.
- Kong, Y. Y., J. M. Deeks, P. R. Axon, and R. P. Carlyon. 2009. Limits of temporal pitch in cochlear implants. *The Journal of the Acoustical Society of America* 125 (3) (Mar): 1649-57.
- Kraus, N., and B. Chandrasekaran. 2010. Music training for the development of auditory skills. *Nature Reviews.Neuroscience* 11 (8) (Aug): 599-605.
- Kraus, N., R. J. Zatorre, and D. L. Strait. 2014. Editors' introduction to hearing research special issue: Music: A window into the hearing brain. *Hearing Research* 308 (Feb): 1.
- Lassaletta, L., A. Castro, M. Bastarrica, R. Perez-Mora, B. Herran, L. Sanz, M. J. de Sarria, and J. Gavilan. 2008. Musical perception and enjoyment in post-lingual patients with cochlear implants. *Acta Otorrinolaringologica Espanola* 59 (5) (May): 228-34.
- Lassaletta, L., A. Castro, M. Bastarrica, R. Perez-Mora, R. Madero, J. De Sarria, and J. Gavilan. 2007. Does music perception have an impact on quality of life following cochlear implantation? *Acta Oto-Laryngologica* 127 (7) (Jul): 682-6.
- Limb, C. J., and A. T. Roy. 2014. Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing Research* 308 (Feb): 13-26.
- Lo, C. Y., C. M. McMahon, V. Looi, and W. F. Thompson. 2015. Melodic contour training and its effect on speech in noise, consonant discrimination, and prosody perception for cochlear implant recipients. *Behavioural Neurology* 2015 : 352869.
- Magne, C., D. Schon, and M. Besson. 2006. Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience* 18 (2) (Feb): 199-211.
- McDermott, H. J. 2004. Music perception with cochlear implants: A review. *Trends in Amplification* 8 (2): 49-82.
- Micheyl, C., K. Delhommeau, X. Perrot, and A. J. Oxenham. 2006. Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research* 219 (1-2) (Sep): 36-47.
- Olszewski, C., K. Gfeller, R. Froman, J. Stordahl, and B. Tomblin. 2005. Familiar melody recognition by children and adults using cochlear implants and normal hearing children. *Cochlear Implants International* 6 (3) (Sep): 123-40.
- Oxenham, A. J. 2008. Pitch perception and auditory stream segregation: Implications for hearing loss and

cochlear implants. *Trends in Amplification* 12 (4) (Dec): 316-31.

- Parbery-Clark, A., E. Skoe, C. Lam, and N. Kraus. 2009. Musician enhancement for speech-in-noise. *Ear and Hearing* 30 (6) (Dec): 653-61.
- Patel, A. D. 2014. Can nonlinguistic musical training change the way the brain processes speech? the expanded OPERA hypothesis. *Hearing Research* 308 (Feb): 98-108.
- Philips, B., B. Vinck, E. De Vel, L. Maes, W. D'Haenens, H. Keppler, and I. Dhooge. 2012. Characteristics and determinants of music appreciation in adult CI users. *European Archives of Oto-Rhino-Laryngology* 269 (3) (Mar): 813-21.
- Ruggles, D. R., R. L. Freyman, and A. J. Oxenham. 2014. Influence of musical training on understanding voiced and whispered speech in noise. *PloS One* 9 (1) (Jan 28): e86980.
- Russell, J. A., and A. Mehrabian. 1977. Evidence for a three-factor theory of emotions. *Journal of Research in Personality* 11 (3): 273-94.
- Salimpoor, V. N., M. Benovoy, G. Longo, J. R. Cooperstock, and R. J. Zatorre. 2009. The rewarding aspects of music listening are related to degree of emotional arousal. *PloS One* 4 (10) (Oct 16): e7487.
- Schon, D., C. Magne, and M. Besson. 2004. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology* 41 (3) (May): 341-9.
- Souza P., and Rosen S. 2009. Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech. *J. Acoust. Soc. Am.* 126, 792–805.
- Strait, D. L., N. Kraus, A. Parbery-Clark, and R. Ashley. 2010. Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research* 261 (1-2) (Mar): 22-9.
- Sucher, C. M., and H. J. Mcdermott. 2009. Bimodal stimulation: Benefits for music perception and sound quality. *Cochlear Implants International* 10 (S1): 96-9.
- Swaminathan, J., C. R. Mason, Ti. M. Streeter, V. Best, G. Kidd Jr, and A. D. Patel. 2015. Musical training, individual differences and the cocktail party problem. *Scientific Reports* 5 .
- Tajudeen, B. A., S. B. Waltzman, D. Jethanamest, and M. A. Svirsky. 2010. Speech perception in congenitally deaf children receiving cochlear implants in the first year of life. *Otology & Neurotology : Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology* 31 (8) (Oct): 1254-60.
- Thompson, W. F., E. G. Schellenberg, and G. Husain. 2004. Decoding speech prosody: Do music lessons help? *Emotion (Washington, D.C.)* 4 (1) (Mar): 46-64.
- Trehub, S. E., T. Vongpaisal, and T. Nakata. 2009. Music in the lives of deaf children with cochlear implants. *Annals of the New York Academy of Sciences* 1169 (Jul): 534-42.
- Van Borsel, J., J. Janssens, and M. De Bodt. 2009. Breathiness as a feminine voice characteristic: A perceptual approach. *Journal of Voice* 23 (3): 291-4.
- Whitmal N. A., Poissant S. F., Freyman R. L., and Helfer K. S. 2007. Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. *J. Acoust. Soc. Am.* 122, 2376–2388.
- Winn, M. B., M. Chatterjee, and W. J. Idsardi. 2011. The perception of phonetic features and acoustic cues by impaired listeners. *The Journal of the Acoustical Society of America* 130 (4): 2374.
- Xin, Luo, Q. J. Fu, and J. J. Galvin 3rd. 2007. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification* 11 (4) (Dec): 301-15.
- Xu, Li, C. S. Thompson, and B. E. Pfingst. 2005. Relative contributions of spectral and temporal cues for phoneme recognition. *The Journal of the Acoustical Society of America* 117 (5): 3255-67.
- Zatorre, R. J. 2013. Predispositions and plasticity in music and speech learning: Neural correlates and implications. *Science (New York, N.Y.)* 342 (6158) (Nov 1): 585-9.

- Zatorre, R. J., and S. R. Baum. 2012. Musical melody and speech intonation: Singing a different tune. *PLoS Biology* 10 (7): e1001372.
- Zatorre, R. J., and V. N. Salimpoor. 2013. From perception to pleasure: Music and its neural substrates. *Proceedings of the National Academy of Sciences of the United States of America* 110 Suppl 2 (Jun 18): 10430-7.
- Zendel, B. R., and C. Alain. 2013. The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *Journal of Cognitive Neuroscience* 25 (4): 503-16.

Summary
Samenvatting



SUMMARY

Cochlear implants (CIs) are prosthetic devices that restore hearing in deaf people who do not benefit from acoustic hearing aids. The perception of speech in quiet environments is for most CI users reasonably good. The perception of other acoustical signals, such as speech in noisy environments or listening to music, is still not satisfying. This thesis focused on the music perception and the potential positive effects of music training on speech and music perception in CI users. First, pitch based speech perception and music perception was subjectively and behaviorally investigated in CI users. Second, the possible beneficial effect of musical training on music and speech perception was investigated.

The music perception and enjoyment of two groups of CI users was investigated using questionnaires. As music can influence quality of life in certain patient groups, the music perception and enjoyment was also related to the quality of life. A large cohort of typical, post-lingually deafened CI users and a group of atypical CI users, the early-deafened, late-implanted group were compared. The early-deafened, late-implanted CI user is deafened during language acquisition, has been hearing deprived for a long time, and is implanted at a later age. The typical CI users is deafened after language acquisition, has a relative short period of auditory deprivation and is implanted at adult age. Early-deafened, late-implanted CI users enjoy music more, rate the quality of the sound of music higher and rate the perception of the elements of music differently than the post-lingually deafened CI users. The sound of music was rated suboptimal in both groups. A better perception of the elements of music was associated with a higher quality of life in the post-lingually deafened CI group; no such correlation was found in the EDLI CI users. The two groups of CI users rate and enjoy music differently, perhaps based on a different acoustical memory of music. Improvement is needed as both groups did rate the quality of music as suboptimal.

To see if the perception of music and its elements is also behaviorally suboptimal, CI users and normal-hearing (NH) listeners were tested for pitch based speech tasks. Voice gender categorization (a female has a higher pitched voice than a male speaker) or the identification of a vocal emotion (happy has a higher pitch than sad) was investigated. Gender categorization was tested, using a female talker that was adjusted via manipulation of two main cues of a voice: the F0, the mean pitch, and the vocal tract length, the distance from the vocal chords to the lips of a talker. Gender categorization was abnormal in CI users in comparison to NH listeners. CI users weighted the F0 of the signal higher versus a lower weighting of the vocal tract length differences, whereas NH listeners weigh both cues more evenly. The results might imply difficulties in daily life situations for gender categorization in CI users. Emotion identification was investigated using a nonce-word and four emotions. Acoustically, three cues available for emotion identification: mean pitch, pitch range and

number of dominant pitches were analyzed. The NH listeners outperformed the CI users. Emotion identification was different in CI users compared to NH listeners. NH listeners utilize the mean pitch, while CI users utilize the pitch range to identify vocal emotions. Perhaps the CI was unable to capture the mean pitch accurately enough for emotion identification. These results indicate pitch-related speech tasks such as gender categorization or emotion identification to be problematic for CI users.

To improve the perception of pitch based speech tasks or music in CI users one could, either try to improve the device via for example signal processing or device design, or one could try to improve patient related factors, such as trying to improve cognitive elements. In this thesis improvements on the patient side were chosen. Musical training has been shown to have benefits for auditory perception in NH listeners. The benefit of musical training, referred to as the 'musician effect', has been shown for music perception (within domain), but even more interesting also for speech perception in noise (transfer effect from music to speech). Areas in which CI users experience difficulties. To see whether a musician effect also exists in CI users or in the degraded CI signal tested via CI simulations, we conducted 3 studies: one subjective and two behavioral.

The first study questioned the musical background in a large cohort of post-lingually deafened CI users. The results indicated that CI users had no or minimal musical background before implantation. No correlations between the musical background and speech perception or with the quality of life were shown. Thus, no musician effect could be shown. It must be noted that the amount of musical training was so minimal in our cohort that perhaps this conclusion may not be drawn.

The second study tested the musician effect behaviorally for gender categorization. NH musicians and non-musicians listened to normal acoustical stimuli and CI-simulations. A musician effect was shown for both the normal and the CI-simulations. The musicians used F0 more than VTL. This could imply a better pitch perception in degraded conditions for musicians. This was a surprising find as the CI users also use the F0 of the signal to categorize the gender of a talker, something that was named abnormal, even more surprising while CI users were shown to have little musical training. An explanation might be that VTL is less reliably conveyed in both CI simulations and for CI users. Therefore, utilizing F0 is advantageous, as it might be the more robust cue in degraded situations.

The third study investigated the musician effect in NH listeners for speech perception, emotion identification and melodic contour identification; identification of a 5-note melodic contour. Again normal acoustical stimuli and CI simulations were used. Results showed only a small transfer effect of musical training on speech for only the CI simulated word in noise condition. No transfer effect for speech perception was shown when listening to normal acoustical stimuli. Musicians did outperform the non-musicians for emotion identification and melodic contour identification for both normal and CI-simulated stimuli. The musician effect became more apparent as the specific task was more pitch based; so a bigger effect for melodic contour identification than for emotion identification was found. Concluding, musicians outperform non-musicians for emotion and melodic contour identification and utilize different cues, more pitch based, than non-musicians. Only a small transfer effect of music to speech perception was shown, in one CI simulated word intelligibility test. A musician effect seems to be apparent when listening to CI simulations.

The last question was if the musician effect could also be shown in real CI users. A feasibility-study with three training groups was conducted. The groups were a musical training, a music therapy, and a control group; the non-musical training. During a six-week period in total 19 CI users were trained weekly. Before and directly after the training period the CI users tested for speech perception in noise, emotion identification and melodic contour identification. Results showed a transfer effect of music therapy on emotion identification and a within domain effect of musical training on melodic contour identification. Subjectively, the CI users in the music therapy group stated that they felt better about their perceptual skills, that they recognized emotions better, and that they began to listen to music more and enjoyed music. These findings shed a possible positive light on the effect of musical training or music therapy on the perception of CI users. This might lead to the inclusion of a music based training or therapy in the rehabilitation of CI users in clinical practice.

In conclusion, this thesis showed that the subjective perception of music in CI users differs per implantee group, but is not satisfying yet. Behaviorally tested gender categorization is abnormal and the emotion identification is impaired in CI users. The musician effect showed possible positive benefits auditory perception in NH listeners. NH musicians outperformed NH non-musicians for CI simulated tasks, more so if the task was more pitch based. A musician effect was found for gender categorization, emotion identification and melodic contour identification. In CI users a possible positive effect of musical training on melodic contour identification and of music therapy on emotion identification was found. To improve the perception of speech, pitch-related speech and music tasks, music training or musical therapy seems promising for the future rehabilitation of CI users.

SAMENVATTING

Een cochleair implantaat (CI) is een chirurgisch geïmplanteerde gehoorprothese voor zeer ernstig slechthorenden, die geen baat hebben bij conventionele hoortoestellen. De meeste CI-gebruikers verstaan spraak in stilte redelijk tot goed. Andere akoestische signalen, zoals spraakverstaan in ruis of het luisteren naar muziek is voor CI-gebruikers uitdagend. Dit promotieonderzoek onderzocht de perceptie van muziek met een CI en de mogelijk positieve invloed van muzikale training op het horen met een CI. In het eerste deel van dit promotieonderzoek is de perceptie van muziek, vocale emoties en het herkennen van het geslacht van een spreker onderzocht in CI-gebruikers. Het tweede deel onderzocht de mogelijk positieve invloed van muzikale training op de waarneming van akoestische signalen, zoals muziek en spraakverstaan in ruis, in normaalhorenden (NH) en CI-gebruikers.

Het genieten van het luisteren naar en het waarnemen van muziek is voor twee groepen CI-gebruikers met vragenlijsten onderzocht. De ene groep is een typische CI-groep: de post-linguaal dove CI-gebruiker; de andere groep is een atypische CI-groep: de vroeg-dove, laat-geïmplanteerde CI-gebruiker. De vroeg-dove, laat-geïmplanteerde CI-gebruiker is ernstig slechthorend of doof geworden tijdens de taalverwerving, maar pas op latere leeftijd, na een langere periode van slechthorendheid, geïmplanteerd. Dit in tegenstelling tot de typische CI-gebruiker, de post-linguaal dove volwassene, die na de taalverwerving als volwassene doof is geworden en geïmplanteerd. De vragenlijsten lieten zien dat de post-linguaal dove CI-gebruiker minder van het geluid van muziek geniet, dan de vroeg-dove, laat-geïmplanteerde CI-gebruiker. Het geluid van muziek was suboptimaal in beide groepen. In de typische groep CI-gebruikers was een betere subjectieve waarneming van de elementen van muziek gerelateerd aan een hogere kwaliteit van leven; deze correlatie was niet aanwezig in de vroeg-dove, laat-geïmplanteerde groep. Wellicht is het verschil tussen beide groepen te verklaren door het verschil in akoestisch geheugen voor muziek. Concluderend lieten de vragenlijsten zien dat de subjectieve perceptie van muziek suboptimaal en verschillend is tussen beide groepen CI-gebruikers en dus verbeterd zou kunnen worden.

De vraag is echter of ook het daadwerkelijk luisteren naar muziek suboptimaal is in CI-gebruikers. Met andere woorden nemen CI-gebruikers ook daadwerkelijk akoestische signalen, zoals muziek, anders of slechter waar dan bijvoorbeeld NH. In dit proefschrift is het herkennen van emoties en het geslacht van een spreker op basis van alleen het geluid onderzocht. Dit zijn beiden taken gebaseerd op het herkennen van de toonhoogte, een belangrijk element van muziek. Emotieherkenning is getest met een nonsens woord, een woord zonder betekenis, dat werd uitgesproken met vier verschillende emoties. CI-gebruikers herkenden de emoties slechter dan de NH. Om na te gaan waarom CI-gebruikers de emoties slechter waarnamen, zijn verschillende akoestische parameters die emotieherkenning en nadruk in emoties kenmerken, zoals de gemiddelde toonhoogte, de range van de toonhoogtes en het aantal dominante toonhoogtes, vergeleken. Hieruit blijkt dat CI-gebruikers de range van de toonhoogte gebruiken voor emotieherkenning, terwijl NH de gemiddelde toonhoogte gebruiken. Mogelijk is de CI-gebruiker onvoldoende in staat om de gemiddelde toonhoogte waar te nemen.

In een tweede studie werd het categoriseren van het geslacht van een spreker, het differentiëren tussen man en vrouw op basis van het stemgeluid alleen, getest. Door het aanpassen van de afstand van de stemplooien tot aan de lippen (VTL) en de toonhoogte van de stem (F0) via een glijdende schaal, kan een vrouwenstem die geleidelijk overgaat in een mannenstem gesimuleerd worden. De resultaten toonden dat CI-gebruikers de vrouwen- en mannenstemmen anders categoriseerden dan NH. Geïmplanteerden gebruikten vrijwel alleen de toonhoogte om een verschil tussen sprekers waar te nemen. NH gebruikten zowel de toonhoogte als de afstand van de stemplooien tot de lippen. Wellicht wordt de VTL , de maat voor de afstand van de stemplooien tot aan de lippen, niet goed waargenomen door CI-gebruikers. De uitkomsten van beide studies tonen aan dat CI-gebruikers waarschijnlijk in het dagelijks leven problemen ervaren bij het herkennen van emoties of het geslacht van een spreker op basis van het geluid alleen. Samengevat toonden deze twee studies dat de identificatie van emoties slechter en anders is in CI-gebruikers, en dat CI-gebruikers het geslacht van een spreker anders categoriseren dan NH. Opnieuw een reden om te onderzoeken of de waarneming van deze akoestische signalen verbeterd zou kunnen worden.

Een mogelijke verbetering van de identificatie van emoties, het geslacht van een spreker of muziek zou kunnen worden gevonden in ofwel het verbeteren van het CI, via bijvoorbeeld betere bewerking en verwerking van het signaal in de processor van de CI, ofwel via het verbeteren van de CI-gebruiker zelf, via bijvoorbeeld training. In dit proefschrift hebben we onderzocht of muzikale training, het zogenoemde 'musicus effect', een verbetering van het horen in CI-gebruikers kan geven. Het musicus effect is het voordeel dat NH-musici hebben bij het waarnemen van bepaalde akoestische stimuli. Zo herkennen musici muziek, maar wellicht interessanter spraakverstaan in ruis beter waar dan niet-musici. De vraag is of dit musicus effect ook bestaat wanneer akoestische signalen met een CI worden waargenomen. Om antwoord te geven op deze vraag hebben we het musicus voordeel in CI-gebruikers en NH onderzocht.

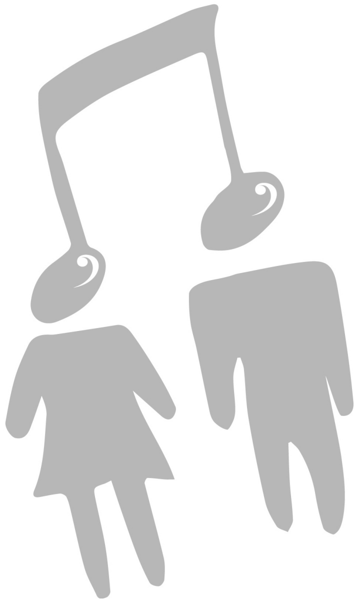
Allereerst is de muzikale training van de CI-gebruiker voor implantatie onderzocht met vragenlijsten. Vervolgens werd gekeken of er een relatie bestaat tussen deze muzikale training en het spraakverstaan en de kwaliteit van leven. CI-gebruikers blijken nauwelijks muzikale training te hebben en er werd geen correlatie aangetoond van de muzikale training met spraakverstaan en de kwaliteit van leven. Wel is het beter waarnemen van de elementen van muziek, zoals toonhoogte en timbre, gerelateerd aan een hogere kwaliteit van leven. Het musicus effect was dus niet duidelijk aanwezig in de geteste CI-gebruikers, hoewel hierbij moet worden aangemerkt dat de CI-gebruikers in onze groep nauwelijks muzikale training hadden.

Om toch te kunnen nagaan of het musicus effect bestaat voor CI's, is een experiment met NH-musici en niet-musici gedaan. In deze studie hoorden de deelnemers normale akoestische stimuli en CI-gesimuleerde stimuli. De deelnemers luisterden net als de CI-gebruikers in de eerdere studies naar emoties, naar het geslacht van een spreker, maar ook naar melodieën en naar spraak in stilte en in ruis. Musici bleken het geslacht van een spreker anders te categoriseren dan niet-musici. Musici gebruiken de toonhoogte (F0) meer dan de afstand van de stemplooien tot de lippen (VTR), terwijl niet-musici beiden gebruiken. Een interessante bevinding omdat ook CI-gebruikers (zonder muzikale training) de toonhoogte gebruikten om het geslacht te categoriseren. Wellicht wordt de afstand van de stemplooien van de lippen niet afdoende doorgegeven via het CI en is dus de toonhoogte het deel van het signaal wat het beste het geslacht categoriseert in CI-geluid. Musici herkenden emoties en melodieën beter dan niet musici voor normale signalen, maar ook voor CI-simulaties. Alle drie taken die deels afhankelijk zijn van een betere herkenning van toonhoogtes. Musici bleken echter maar in één conditie van de CI-simulaties beter in het spraakverstaan in ruis. Kortom, musici zijn ook in situaties waarin het akoestische signaal minder rijk is ofwel CI-gesimuleerd, in staat om zowel spraak gerelateerde taken, zoals emotieherkenning, als muziek beter waar te nemen dan niet-musici. Het musicus effect lijkt dus wel te bestaan bij het luisteren naar CI-simulaties en was duidelijker aanwezig naarmate de taak meer op het herkennen van de toonhoogte gebaseerd was.

Afsluiten hebben we een korte trainingsstudie met CI-gebruikers verricht om te onderzoeken muzikale training het horen verbeterd. Negentien CI-gebruikers werden verdeeld over drie trainingsgroepen: muzikale training, muziektherapie en niet-muzikale training. Individueel of in groepen hebben ze vervolgens zes weken lang, twee uur per sessie getraind. De resultaten lieten zien dat muzikale training het waarnemen van melodieën verbeterd en dat muziektherapie de emotieherkenning verbeterd. Het spraakverstaan verbeterde niet in de drie trainingsgroepen. Subjectief gaven de CI-gebruikers in de muziektherapiegroep aan dat ze meer van muziek genoten en dat ze vonden dat ze muziek beter waar konden nemen na de therapie. Hoewel moet worden aangemerkt dat dit een kleine studie was uitgevoerd binnen een kort tijdsbestek, bieden deze resultaten hoop voor het mogelijk positieve effect van muzikale training of muziektherapie op de waarneming van melodieën en emoties in CI-gebruikers.

De resultaten van dit proefschrift laten zien dat het waarnemen van muziek door CI-gebruikers kan worden verbeterd. CI-gebruikers hebben een abnormale waarneming van het geslacht van een spreker en een verminderde herkenning van emoties. In NH werd een musicus effect aangetoond luisterend naar CI-simulaties voor emotieherkenning, categorisatie van het geslacht van een spreker en het herkennen van melodieën. Het musicus effect is waarschijnlijk gebaseerd op een betere waarneming van toonhoogtes. Hoewel er tussen de subjectieve muzikale training van CI-gebruikers en de waarneming van spraak geen relatie is gevonden, zou muzikale training of muziektherapie in de revalidatie kunnen worden toegepast om het horen met een CI te optimaliseren, omdat een korte trainingsperiode al positieve resultaten liet zien. De revalidatie zou zich meer kunnen richten op het trainen van de muziek.

Dankwoord



Dit werk zou niet tot stand zijn gekomen zonder het enthousiasme, doorzettingsvermogen en de hulp van velen.

Allereerst wil ik alle CI-gebruikers, musici en niet-musici die aan de studies hebben meegewerkt van harte danken. Zonder het enthousiasme van alle deelnemers zou de wereld niet een klein beetje wijzer zijn geworden. Het was een feest om zowel in de stille kamer, de faculteit, als in het voormalig CI-huis in Haren met een ieder van jullie te mogen samenwerken. Het was een fantastisch project. Hopelijk krijgt het een waardig vervolg!

Geachte professor Başkent, dear Deniz, you taught me the first steps in the research world and lead the music and CI team with never exhausting enthusiasm and positivism (and budget) to the borders of our knowledge. Thank you for your leading hand towards this finished project, the conversations about all topics in life in Groningen or somewhere on this planet and your hospitality. Even though it is not your quote, I feel it marks the feeling of this project and your research group: Science is fun!

Beste Rolien, dank voor al het vertrouwen dat je in me hebt gehad en alle steun die je me hebt geboden op alle vlakken en overal ter wereld. Jouw begeleiding in het onderzoek en tijdens de opleiding is me zeer veel waard. Vanuit jouw enthousiasme en idee als CI-dokter en onderzoeker is mijn muziekproject gegroeid en zal het hopelijk nog blijven groeien!

Geachte professor Van der Laan, dank voor alle onderzoeksmogelijkheden die mij zijn geboden en hopelijk in de toekomst geboden blijven binnen de opleiding tot KNO-arts. Ondanks de soms moeilijke gesprekken die we moesten voeren, heb ik me altijd gesteund gevoeld.

Beste Bert Maat, dank voor de tijd die je in de CI-database en mij hebt gestoken. Zonder jouw uitleg zou de CI en zou Access voor mij een groot raadsel zijn gebleven.

Dear John Galvin, thanks to you I will never get rid of my craving feeling for a Margarita as soon as I spot an ocean, but moreover thank you for the co-work and the English lessons, mostly in red, you taught me.

Dear Etienne Gaudrain, little red riding hood will never be the same again. With or without a CI the fairytale of life is sometimes a hard one, but I have faith in you and in the happy ever after. I really appreciate our friendship and co-work on the gender categorization.

Graag wil ik ook de leescommissie, prof. dr. A. Faulkner, prof. dr. ir. J.H.M. Frijns en prof. dr. ir. N.M. Maurits, bedanken voor de tijd die jullie genomen hebben voor de beoordeling en goedkeuring van het manuscript. Thank you very much for the time and effort invested in the thesis.

Carmen, Roy, Jooske, Han, Karin, Joeri, Aline, Saar, Angelique, Gerda, Steven en Dicky: dank voor de samenwerking in dit onderzoek, het enthousiasme en de discussies!

Loes
Lieve Loes, al badmintonnend hebben we heel Nederland gezien, vele partijen verloren en gewonnen, maar de grootste prijs die ik heb gewonnen in mijn badmintoncarrière ben jij. Samen met Jasper en jullie drie kinderen zijn jullie en ben jij de vriendin/ de rots in de branding die iedereen zich wenst. Als paranimf zal je mij vandaag begeleiden, maar je weet het: ik hoop ook altijd voor jou klaar te staan.

Ellen
Roomie! Wat ben je toch een topper. In de goede tijden hebben we de grootste lol en in moeilijke tijden sta je voor me klaar. Naast collega's zijn we goede vrienden geworden en is onze vriendschap uitgegroeid tot een zeer waardevolle! Ik ben dan ook heel erg blij dat jij vandaag eens een Poosje naast mij staat.  We can do it!

Assistenten en bazen KNO
Wat fijn om op vrijdag tijdens de staf met iedereen te lachen, taartjes te eten en af en toe een oogje dicht te knijpen, om nadien in de borrelkamer de opleiding en het leven eens te evalueren. Ik voel me een deel van een fijn en leergierig team!

Hilde en Hilde: dank voor alle lekkere dinertjes en alle gesprekken over de geneeskunst en de wetenschap!

DDNL – Miriam, Mynke en Sarah
Lieve dames Neerlandici, als ik zeg schuif even op, zeggen jullie… Zo begon het ooit en zo is het gebleven. De sloten kovvie en borden eten die we met elkaar op verschillende locaties hebben gedeeld en zullen delen, maken me iedere keer weer gelukkig. En zoals echte vriendschap betaamt kunnen we met elkaar huilen en lachen, maar bovenal ook praten en zwijgend naast elkaar zitten en weten dat het goed is zo. Dank voor jullie steun en vertrouwen.

Mama
Allerliefste mama, wat hebben we een roerige jaren gehad, maar we bouwen aan een nieuw hoofdstuk voor een stralende toekomst. Dank voor alle mogelijkheden die jullie mij hebben gegeven en het vertrouwen dat jullie altijd in me gehad hebben. Gelukkig is wat je ooit in mijn poëziealbum schreef, uitgekomen: hopelijk kunnen we later vriendinnen zijn.

Liebe Jan, you make my life complete and I want to complete my life with you, wherever whenever. Du bist mein Lieblingsmensch, wie Kaffee auf Hawaii.

CURRICULUM VITAE

Christina Fuller werd geboren op 2 juni 1985 in het Drentse Borger. Na afronding van het gymnasium aan het *Esdal College* in Emmen, startte ze in 2003 de studie Nederlandse taal en cultuur aan de *Rijksuniversiteit Groningen*. In 2004 begon ze hiernaast aan de studie geneeskunde aan dezelfde universiteit. Na het afronden van haar studie Nederlands in 2008 met de master thesis 'De inleidende gedichten van Jacob Cats voor Johan van Beverwijck in het eerste deel van de Schat der Gesontheyt', startte zij haar co-schappen, in het *Universitair Medisch Centrum Groningen* en het *Deventer Ziekenhuis*. In 2011 rondde zij de studie geneeskunde succesvol af, nadat zij zich in het laatste jaar van de studie had gericht op de keel-, neus- en oorheelkunde. Haar master thesis richtte zich reeds op de muziekbeleving en de kwaliteit van leven van cochleair implantaat gebruikers. Aansluitend startte zij haar promotietraject in 2011 op de afdeling KNO van het UMCG onder de leiding van prof. dr D. Başkent, dr. R.H. Free en prof. dr. B.F.A.M. van der Laan. Vanaf november 2013 is zij in opleiding tot keel-, neus- en oorarts in het UMCG met als opleider prof. dr. B.F.A.M. van der Laan en in het Medisch Centrum Leeuwarden onder dr. H. van den Berge.