Measure and model of vocal-tract length discrimination in cochlear implants

Etienne Gaudrain, Lucas Stam, Deniz Başkent
University of Groningen, University Medical Center Groningen
Department of Otorhinolaryngology / Head and Neck Surgery
Groningen, The Netherlands
e.p.c.gaudrain@umcg.nl

Abstract— **Voice discrimination is crucial to selectively listen to a particular talker in a crowded environment. In normal-hearing listeners, it strongly relies on the perception of two dimensions: the fundamental frequency and the vocal-tract length. Yet, very little is known about the perception of the latter in cochlear implants. The present study reports discrimination thresholds for vocal-tract length in normal-hearing listeners and cochlear-implant users. The behavioral results were then used to determine the effective spectral resolution in a model of electric hearing: effective resolution in the implant was found to be poorer than previously suggested by psychophysical measurements. Such a model could be used for clinical purposes, or to facilitate the development of new strategies.**

Keywords—cochlear implant; voice; vocal-tract length; modeling; spectral resolution

I. INTRODUCTION

In a crowded cocktail party, understanding speech produced by one talker is closely linked to being able to selectively listen to this specific talker among other speakers. The difficulty to understand a talker when another talker is speaking at the same time is directly related to how different the voices of the two speakers are [1]. This difference can be quantized along two anatomically related dimensions: the fundamental frequency (F0), which is associated with pitch perception, and the vocal-tract length (VTL), which is associated with the size of the speaker [2]. In the acoustic signal, F0 determines the scale of the harmonic structure, while VTL constrains the scale of the spectral envelope along the frequency axis. Both dimensions were shown to provide similar advantage for speech-on-speech perception [3].

While F0 perception has been extensively studied in cochlear implants (CI) [4], little is known about VTL perception. A recent study showed that, unlike normal-hearing (NH) listeners who use both F0 and VTL to categorize the gender of a voice, CI listeners rely exclusively on F0 cues, indicating that VTL cues are not perceived [5]. Further examination suggests that the poor spectral resolution available through the implant is the principal suspect for this lack of sensitivity to VTL differences [6]. This deficit not only makes voice gender identification difficult, it also suggests that CI listeners may have access to only one voice cue instead of two to separate concurrent voices, and derive a benefit in speech intelligibility.

In the study reported here, we sought to characterize VTL perception in CI users by measuring the just-noticeable-difference (jnd) along that dimension. The obtained VTL jnds, which represent the smallest VTL values that can be detected, were compared with those obtained in NH listeners.

Current spread in the cochlea is largely responsible for the poor spectral resolution available through implants. To examine how current spread affects VTL discrimination, models of acoustic and electrical hearing based on the Auditory Image Model [7], [8] were used. This approach shows that typical values of current spread are predictive of the VTL jnds observed in CI users.

II. BEHAVIORAL MEASURE

A. Methods

The VTL jnds were obtained using a three interval, three alternative forced choice (3I-3AFC) adaptive procedure producing a threshold corresponding to the 70.7% point of the psychometric function [9]. In each trial, three consonant-vowel syllables were randomly selected from a list of 61 Dutch syllables recorded from a female speaker. The syllables were processed with STRAIGHT [10] to effect changes in VTL relative to those of the original speaker. With this method, two versions of the syllable triplet were created: one where the syllables were resynthesized with the original voice parameters (standard triplet), and one where the syllables were resynthesized with altered VTL (test triplet). Two identical instances of the standard triplet and one instance of the test triplet were then presented in a random order and the participant was instructed to identify the triplet that differed from the other two. On the initial trial, the VTL difference was 12 semitones (st), i.e. the processed VTL was twice as long as the original one. When the participants had two successive correct responses, the VTL difference was reduced by 2 st. On each incorrect response, the VTL difference increased by that same step-size. When the difference became smaller than twice the step-size, the step-size was reduced by a factor $\sqrt{2}$. The adaptive procedure stopped after 8 reversals and the jnd was calculated as the average of the VTL differences at the last 5 reversals. The procedure was repeated three times for each participant and the obtained jnds were averaged.

The auditory stimuli were presented at a level of 65 dB SPL either over headphones (Sennheiser HD 650) in a sound-treated booth for the NH participants, or over a loudspeaker (Tannoy)

The study was supported by a NWO/ZonMW-VIDI grant (016.096.397), a Rosalind Franklin Fellowship from the University of Groningen, and funds from Heinsius Houbolt Foundation. The study is part of the research program of the Otorhinolaryngology Department of University Medical Center Groningen: Healthy Aging and Communication.

in an anechoic room for the CI participants. In both setups the sounds were delivered by an AudioFire4 soundcard (Echo) connected to a D/A converter (DA10, Lavry) through an S/PDIF link.

Sixteen NH listeners were recruited to take part in the experiment. They were aged 19 to 63 and all had audiometric thresholds ≤ 20 dB HL at octave frequencies between 500 and 4000 Hz. Six CI users, aged 48 to 69, took part in the experiment. Five were users of the Cochlear Nucleus CI24R device while one used the Advanced Bionics HiRes 90K device. All the participants provided signed informed consent prior to data collection. The experiment was approved by the ethics committee of the University Medical Center Groningen (METc 2012.392). The volunteers received an hourly wage for their participation.

B. Results and discussion

Average results for the NH listeners and individual results for the CI listeners are shown in Fig. 1. The average VTL jnd is 1.6 st for NH listeners and 5.6 st for CI listeners [$t(5.4)=5.4$, $p=0.002$]. All the CI users but one had jnds greater than 4 st. This means that they would not be able to detect the typical difference in VTL between male and female speakers (below 4 st [11]–[13]).

Detecting a change in VTL requires being able to detect a consistent shift of all the peaks in the spectral envelope. When examining the electrical output of the implant, it is clear that such a shift is indeed conveyed. Frequency channels in CIs are typically separated by 2.0 to 3.5 st. Therefore the typical VTL difference between male and female speakers results in a shift of the electrical excitation pattern of about one whole electrode. The fact that the CI users cannot detect such a shift suggests that the electrical stimulation pattern is not accurately transmitted to the neurons. When electrodes are excited in monopolar mode (i.e. the current is injected in the cochlea and returned through an electrode located outside the cochlea), as it is the case in all implants used clinically, the current delivered by the electrode spreads in the cochlear fluid, resulting in a (spectrally) smeared neural activity pattern. The next section describes a model of electrical stimulation and details the effect of current spread in the cochlea on VTL jnds.

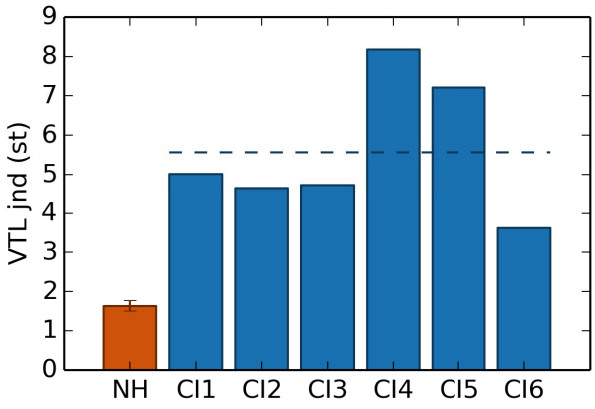

Fig. 1. Average VTL jnd for NH listeners and individual VTL jnds for CI listeners. The error bar for the NH data is the standard error of the mean. The dashed line represents the average VTL jnd for the CI listeners.

III. Model for Electric and Acoustic Hearing

A. General structure

The models described here are based on the Auditory Image Model [7], [8]. From a sound recording, this model produces a two-dimensional image of the neural activity probability (NAP), as a function of time and frequency (or more exactly, place along the tonotopically organized cochlear partition). A *spectral profile* can be obtained by averaging the NAP over time. The VTL jnds are then predicted by calculating the Euclidian distance between the spectral profile of an utterance and that of the VTL-shifted version of this same utterance. This distance is then normalized relative to the standard deviation of the reference spectral profile, across frequencies. This normalized Euclidian distance represents the perceptual distance between two sounds, and is noted D in what follows.

In the first step, the acoustic version of the model (further described below), D was computed for all 61 syllables and for VTL shifts of 1.6 st (the VTL jnd found for NH listeners), and averaged across all syllables. The value of D corresponding to this jnd, D_t, is defined as the perceptual distance necessary for detection of a difference in VTL.

In the second step, a CI version of the model was developed and used to compute values of D for VTL differences of 5.6 and 7 st. The model, described below, has a parameter, λ, reflecting the amount of current spread that happens in the cochlea. The perceptual distance was also calculated for a number of values of λ in order to find the value that would produce a D equal to the jnd definition, D_t.

B. Acoustic model

The model used, implemented in Matlab, is available from http://code.soundsoftware.ac.uk/projects/aimmat. AIM-Mat is composed of various modules (see Fig. 3). For the NH computations, we used the following modules: (1) pre-cochlear processing, 'gm2002' [14]; (2) basilar membrane motion, 'gtfb'; (3) neural activity probability, 'hcl' (half-wave rectification, log-compression, low-pass filtering).

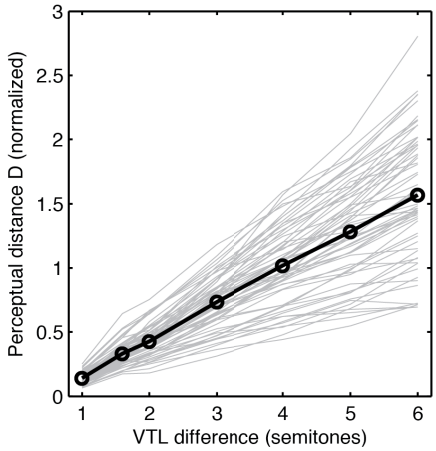

Fig. 2. Acoustic model – Perceptual distance D from the 61 syllables (gray lines), and average (black thick line), as a function of VTL difference.

Fig. 2 shows the calculated values of D for the 61 syllables as well as the average. The relationship between VTL difference and perceptual distance seems quasi-linear. At the NH jnd, for a VTL difference of 1.6 st, the perceptual distance D_t is 0.33 (as a proportion of the standard deviation of the spectral profile).

C. Cochlear implant model

This model, called AIM-CI, has a similar modular structure to AIM (see Fig. 3). The initial pre-processing module is replaced by an implant model producing a map of current levels at each electrode over time. In the current implementation, the Nucleus Matlab Toolbox (Cochlear Ltd.) was used. The following module simulates how the current travels from the electrode to the neurons. In that step, the current spreads across the cochlea according to an exponential decay function [15]–[17] with a spread width λ:

$$I(x) = I_0 \exp\left(\frac{|x-x_e|}{\lambda}\right) \quad (1)$$

Equation (1) gives the current at position x in the cochlea when the electrode located in x_e is activated.

From this time-place current map at the neuron, a neural activity probability map is calculated using a modified version of [18] that is sensitive to pulse shape asymmetries [19].

Typical values for λ reported in the literature range roughly from 2.2 to 4.3 mm, with an average of 3.1 mm [16]. In Fig. 4, the perceptual distance calculated from the CI model is shown for a VTL difference of 5.6 st, and for two values of λ. The threshold perceptual distance D_t was reached for λ=7.2 mm, i.e. more than twice the assumed value.

IV. Conclusion

CI users are less sensitive to VTL differences than NH listeners. The lack of sensitivity can be explained and simulated by the amount of current spread in the cochlea, which results in decreased frequency selectivity compared to NH listeners. Assuming that current spread is the only cause of loss of sensitivity, the VTL jnd measure could provide a new tool to measure effective spectral resolution, which has functional relevance for speech perception. Such a tool could be highly useful in clinical context to optimize the fitting of the implants, or to develop new speech encoding strategies.

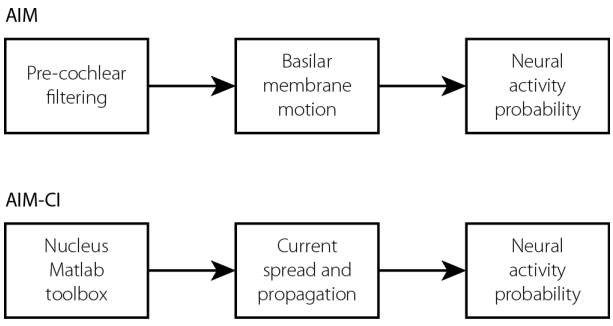

Fig. 3. Block diagram of the AIM (*top*) and of AIM-CI (*bottom*).

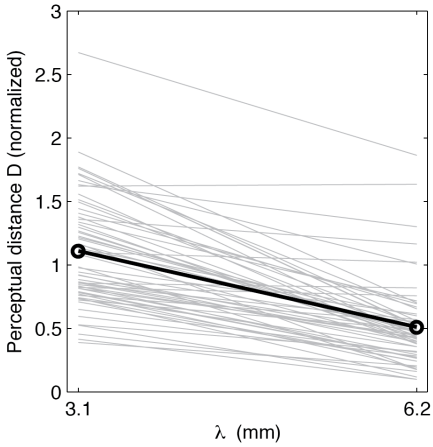

Fig. 4. Cochlear implant model – Perceptual distance D from the 61 syllables (gray lines), and average (black thick line), as a function of current spread width λ, for a VTL difference of 5.6 st.

However it should be noted that other aspects of cochlear implant stimulation than current spread and spectral resolution could also affect VTL sensitivity. For instance, the specific choice of channel boundaries in the frequency allocation map could result in more or less distortion of the VTL cue independently of spectral resolution. If the spectral information were distorted, a VTL difference would not result in a consistent shift of all spectral peaks along the frequency axis. This hypothesis could be tested in a subsequent study by comparing syllables with higher formants to syllables with lower formants.

Acknowledgment

The authors thank the volunteers for taking part in this study, Esmée van de Veen, Floor Burgerhof and Julia Verbist for helping in the collection of the data, as well as Olivier Macherey for multiple advices regarding the CI model.

References

[1] D. S. Brungart, "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.*, vol. 109, no. 3, pp. 1101–1109, Mar. 2001.

[2] W. T. Fitch and J. Giedd, "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1511–1522, 1999.

[3] C. J. Darwin, D. S. Brungart, and B. D. Simpson, "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," *J. Acoust. Soc. Am.*, vol. 114, no. 5, pp. 2913–2922, Nov. 2003.

[4] B. C. J. Moore and R. P. Carlyon, "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch: neural coding and perception*, C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper, Eds. New-York, NY: Springer/Birkhäuser, 2005, pp. 234–277.

[5] C. Fuller, E. Gaudrain, J. Clarke, J. J. Galvin, Q.-J. Fu, R. Free, and D. Başkent, "Gender categorization is abnormal in cochlear-implant users," *J. Assoc. Res. Otolaryngol.*, in revision.

[6] E. Gaudrain and D. Başkent, "Factors limiting vocal-tract length perception in cochlear implants," presented at the 37th Annual Mid-winter Meeting of the Association for Research in Otolaryngology, San Diego, CA, USA, 2014.

[7] S. Bleeck, T. Ives, and R. D. Patterson, "Aim-mat: the auditory image model in MATLAB," *Acta Acust. united Ac.*, vol. 90, pp. 781–787, 2004.

[8] R. D. Patterson, M. H. Allerhand, and C. Giguere, "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Am.*, vol. 98, no. 4, pp. 1890–1894, Oct. 1995.

[9] H. Levitt, "Transformed Up-Down Methods in Psychoacoustics," *J. Acoust. Soc. Am.*, vol. 49, no. 2B, pp. 467–477, Feb. 1971.

[10] H. Kawahara and T. Irino, "Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation," in *Speech separation by humans and machines*, P. L. Divenyi, Ed. Massachusetts: Kluwer Academic, 2004, pp. 167–180.

[11] G. E. Peterson and H. L. Barney, "Control Methods Used in a Study of the Vowels," *J. Acoust. Soc. Am.*, vol. 24, no. 2, pp. 175–184, Mar. 1952.

[12] R. E. Turner, T. C. Walters, J. J. M. Monaghan, and R. D. Patterson, "A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data," *J. Acoust. Soc. Am.*, vol. 125, no. 4, pp. 2374–2386, Apr. 2009.

[13] G. C. M. Fant, *Acoustic Theory of Speech Production*. The Hague: Mouton, 1970.

[14] B. R. Glasberg and B. C. J. Moore, "A Model of Loudness Applicable to Time-Varying Sounds," *J. Audio Eng. Soc.*, vol. 50, no. 5, pp. 331–342, 2002.

[15] R. C. Black and G. M. Clark, "Differential electrical excitation of the auditory nerve," *J. Acoust. Soc. Am.*, vol. 67, no. 3, pp. 868–874, Mar. 1980.

[16] M. Bingabr, B. Espinoza-Varas, and P. C. Loizou, "Simulating the effect of spread of excitation in cochlear implants," *Hear. Res.*, vol. 241, no. 1–2, pp. 73–79, Jul. 2008.

[17] J. Laneau, J. Wouters, and M. Moonen, "Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees," *J. Acoust. Soc. Am.*, vol. 116, no. 6, pp. 3606–3619, Dec. 2004.

[18] S. Fredelake and V. Hohmann, "Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation," *Hear. Res.*, vol. 287, no. 1–2, pp. 76–90, May 2012.

[19] O. Macherey, A. van Wieringen, R. P. Carlyon, J. M. Deeks, and J. Wouters, "Asymmetric pulses in cochlear implants: effects of pulse shape, polarity, and rate," *J. Assoc. Res. Otolaryngol.*, vol. 7, no. 3, pp. 253–266, Sep. 2006.