Susceptibility to interference by music and speech maskers in middle-aged adults

Deniz Başkent[a)]
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
d.baskent@umcg.nl

Suzanne van Engelshoven
Department of Biomedical Engineering, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
s.van.engelshoven@student.rug.nl

John J. Galvin III[b)]
Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
jgalvin@mednet.ucla.edu

Abstract: Older listeners commonly complain about difficulty in understanding speech in noise. Previous studies have shown an age effect for both speech and steady noise maskers, and it is largest for speech maskers. In the present study, speech reception thresholds (SRTs) measured with competing speech, music, and steady noise maskers significantly differed between young (19 to 26 years) and middle-aged (51 to 63 years) adults. SRT differences ranged from 2.1 dB for competing speech, 0.4–1.6 dB for music maskers, and 0.8 dB for steady noise. The data suggest that aging effects are already evident in middle-aged adults without significant hearing impairment.

PACS numbers: 43.71.Lz, 43.71.Sy, 43.66.Dc [DOS]
Date Received: October 23, 2013 **Date Accepted:** December 17, 2013

1. Introduction

Music can have a great effect on mood and emotions, usually positive and pleasurable (Blood and Zatorre, 2001), to the degree that music is used for therapeutic purposes (Hanser and Thompson, 1994). Restaurants and other public places use background music to help customers feel more relaxed (Milliman, 1986). However, background music can interfere with speech understanding and hinder communication (Ekström and Borg, 2011; Eskridge *et al.*, 2012; Gfeller *et al.*, 2012). This could be even more debilitating for older people, who seem to be more sensitive to interference from background sounds in their speech understanding (Bergman *et al.*, 1976; Tun *et al.*, 2002).

Music as a background signal may interfere with speech in several ways (Russo and Pichora-Fuller, 2008; Shi and Law, 2010; Eskridge *et al.*, 2012; Gfeller *et al.*, 2012). Music may mask target speech due to overlap in the neural representations of each signal, i.e., "energetic masking" (Durlach *et al.*, 2003). Music may also produce some degree of "informational masking" due to similarities in some physical properties such as the temporal envelope (Brungart, 2001; Shi and Law, 2010) or due

[a)]Author to whom correspondence should be addressed. Also at: Graduate School of Medical Sciences, Research School of Behavioral and Cognitive Neurosciences, University of Groningen, Groningen, The Netherlands.
[b)]Also at: Department of Head and Neck Surgery, University of California Los Angeles, David Geffen School of Medicine, Los Angeles, CA 90095.

to the semantic content of masker (e.g., meaningful lyrics; Tun *et al.*, 2002). The effects of energetic and informational masking may interact with the age of a listener. Age-related hearing loss, for example, due to widened auditory filters (Başkent, 2006), may increase energetic masking, making background music a more effective masker for typical older listeners than for young listeners (Ekström and Borg, 2011). However, even in normal-hearing individuals where energetic masking effects would be expected to be minimal, the effectiveness of speech and noise maskers has been shown to interact with listener age. Masking by a background talker or multi-talker babble noise on target speech was greater for older than for younger normal-hearing individuals, and the age effect was greater with these maskers than with the steady noise masker (Tun *et al.*, 2002; Rajan and Cainer, 2008). This suggests that informational masking may play an important role in the difficulties understanding speech in noise often experienced by older listeners. Age-related changes in cognitive functioning may additionally contribute to older listeners' poorer speech understanding in noise (Pichora-Fuller *et al.*, 1995; Helfer and Freyman, 2008). Music and speech can share similar spectro-temporal properties, and music can have dominant and dynamic spectral or temporal content, and further, may contain lyrics or emotional meaning. Therefore, music may produce some degree of both energetic and informational masking on target speech.

In the present study, we hypothesized that music maskers would produce greater interference for older listeners than for younger listeners, even for middle-aged listeners with minimal age-related hearing loss, as age-related cognitive changes can start by middle age (Park *et al.*, 2002; Humes *et al.*, 2006). Further, we hypothesized that age effects would be greatest for a speech masker and lowest for a steady noise masker, due to minimal energetic masking effects, and with differences in music maskers falling somewhere in between. To test these hypotheses, we measured older and younger NH listeners' understanding of meaningful sentences presented in competing speech (sentences spoken by the opposite gender talker), steady noise, and four music selections that represented a range of musical qualities (e.g., vocal music with lyrics, instrumental music without lyrics, slow or fast tempo, male or female singer, simple or complex melodies, contemporary or classical genres, etc.). To balance for potential familiarity effects (Russo and Pichora-Fuller, 2008), only very popular and well-known music pieces were selected. To focus on the effects of aging, older participants were selected to be as normal hearing as possible, and audibility was equated between the two groups by low-pass filtering stimuli.

2. Methods

2.1 Participants

Thirteen young (7 women, 6 men; 19–26 years, mean age 23.2 years) and 12 middle-aged (8 women, 4 men; 51–63 years, mean age 58.5 years) normal hearing, native Dutch speakers participated in the study. Exclusion criteria for normal hearing were defined as (1) the pure tone average of thresholds at all test frequencies (250 Hz–8 kHz) for one ear was >20 dB hearing level (HL), (2) a threshold at a single frequency was $>$ 50 dB HL, (3) there was $>$ 10 dB HL difference between thresholds in two ears for at least two frequencies (Stephens, 1996). Five middle-aged people who self-reported to have normal hearing were excluded according to these criteria. The average hearing thresholds at the audiometric frequencies between 250 Hz and 2 kHz were almost identical between the two groups, and differed by 13 dB at 4 kHz (6 dB HL and 19 dB HL for young and older groups, respectively). One young and three older participants were raised in bilingual settings (Dutch/Frisian, and Dutch/Spanish). To characterize music experience, participants were asked two questions: (1) Do you play music? and (2) Do you have perfect pitch? Two young and five older participants reported playing music, and one older participant reported having perfect pitch. Baseline speech intelligibility tests with similar target speech as the ones used in the experiment, but with no masking, were used as additional inclusion criteria. All participants performed well on these tests.

2.2 Stimuli

To equate for the small difference observed in high-frequency thresholds between the two participant groups, all target and masker stimuli were low-pass filtered at 4 kHz, using a Butterworth filter (eighth order; slope 48 dB/octave), thus maintaining equal audibility for both groups. All speech and masker signals were down-sampled to 22 050 Hz, 16 bits. Signal processing was performed using MATLAB.

Target speech consisted of meaningful everyday Dutch sentences, produced by a female talker and digitally recorded at a sampling rate of 44.1 kHz (Versfeld *et al.*, 2000). The stimulus set consisted of 39 sentence lists; each list contained 13 sentences, and each sentence contained 4 to 9 words. The duration of the target sentences ranged from 1.30–2.69 s. List 1 was used for training, and lists 2–39 were used for testing.

Maskers included steady noise, competing speech, and music. During testing, short excepts were randomly selected from within the longer masker files; excerpts were 1 s longer than the target sentence, allowing for a 500 ms onset and offset relative to the target speech.

The steady noise masker was spectrally shaped to match the long-term spectrum of the sentences produced by the male masker talker (Versfeld *et al.*, 2000). Speech maskers were created using sentences that were similar to the target sentences in structure, but were produced by a single male talker (Versfeld *et al.*, 2000). For testing, four speech maskers were created by concatenating sentences from lists 1–9, lists 10–18, lists 19–27, and lists 28–36, resulting in maskers that were 3–4 min in total duration, each containing 117 sentences. The long speech maskers were created to avoid using the same sentence as a masker, as well as to make masker longer than the target sentence. During data collection, one of the four masker files was randomly selected for each participant. For training, a similar long masker was created by concatenating sentences from lists 37–39.

Music maskers consisted of excerpts from one of four different musical pieces: "No. 5 in F sharp major, op. 15 no. 2" by Maria João Pires (original by Frédéric Chopin), "The Cello Song" by the Piano Guys (original by Bach), "Vlieg met me mee" by Trijntje Oosterhuis, and "Dromen zijn bedrog" by Marco Borsato, hereafter referred to as "Chopin," "Cello," "Trijntje," and "Borsato," respectively. All music materials were originally of CD quality, ripped to a PC as .wav files while conserving the highest quality. The stereo music samples were edited using AUDACITY 2.0.2. Music samples were visually inspected and excerpts were selected to maintain a somewhat regular temporal envelope (i.e., no extended periods of high or low amplitude) with a length of 20–30 s and then converted to mono. Chopin contained only one instrument (piano), and had a relatively slow tempo. Cello contained seven instruments (all cellos), and had a relatively fast tempo. Trijntje contained an acoustic guitar with a female singer, and had a relatively slow tempo. Borsato was the most complex piece, containing multiple instruments, a male singer, and a relatively fast tempo. Figure 1 shows temporal waveforms and spectral envelopes for the target speech and masker exemplars, which are provided in Mm. 1. Participants were all familiar with Borsato piece, mostly familiar with Trijntje and Cello, and least familiar with Chopin. The young group reported liking Cello most, then Borsato, then equally Trijntje/Chopin. The older group reported liking Cello/Chopin most, then Borsato, then Trijntje.

Mm. 1. Audio samples of the waveforms shown in Fig. 1, order from the top to bottom.

Due to the adaptive procedures in which the intensity of the masker stimuli varied, it was necessary to maintain sufficient headroom to avoid clipping when the signal and the masker were mixed, especially low signal-to-noise ratios (SNRs). Therefore, all test materials were normalized in intensity to an RMS value of −45 dB FS or lower. This provided enough headroom for testing up to a maximum SNR of −20 dB for all conditions, and, after amplification, limited the intensity of the mixed masker and target to 80 dBA. This prevented uncomfortably loud levels while keeping the noise floor of speech or masker sufficiently low.

2.3 Procedure

First, participants were trained with similar speech and masker materials as used in the experiment. During training, feedback was provided after the subject's response by

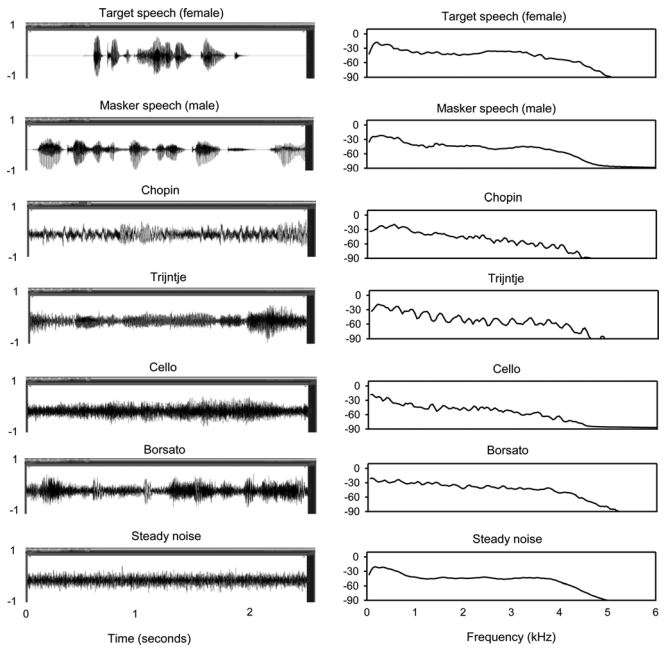

Fig. 1. Examples of temporal waveforms (left column) and spectral envelopes (right column) of experimental stimuli. A target sentence spoken by the female talker is shown in the top row; note that the onset is 500 ms after the onset of the maskers (shown in the bottom six rows) and the offset is 500 ms before the offset of the maskers. All example stimuli have been normalized to have the same RMS amplitude for illustrative purposes. The *y* axis of the temporal waveforms shows the normalized amplitude, and the *x* axis shows time. The long-term spectral envelope was calculated across the entire signal. The *y* axis of the spectral envelope plots shows amplitude re 0 dB FS, and the *x* axis shows frequency in Hz.

playing back the combined target speech and masker and displaying the text of the target sentence on the monitor, similar to Fu and Galvin (2007) and Benard and Başkent (2013). To make sure that every participant became familiar with the music pieces used in the experiment, all four pieces were presented to the participant prior to the experiment, during which participants listened to a longer presentation of the song that included the excerpts that would be used in the experiment.

An adaptive procedure similar to Eskridge *et al.* (2012) was used to measure the masked speech reception threshold (SRT), defined as the SNR required to produce 50% correct word-in-sentence recognition. During testing, a stimulus (masker + target) was presented at the target SNR, and the participant repeated as many words produced by the female target talker as possible. The target speech was presented at 60 dBA and the masker level was adjusted according to subject response. If the subject repeated 50% or more of the words in the target sentence correctly, the masker level was increased by 2 dB. If the subject repeated fewer than 50% of the words in the target sentence correctly, the masker level was reduced by 2 dB. The initial SNR was −4 dB, and the maximum SNR was limited to −20 dB, as explained above. For every trial, the mean SRT over the final 10 out of 13 reversals was measured, and the mean SRT for each subject was calculated from two trials for each condition.

Testing was conducted in a sound-treated booth using custom software (iSTAR; provided by the Emily Shannon Fu Foundation). During each trial, a test sentence was randomly selected from the target sentence list and no sentence was repeated. A masker excerpt was randomly selected from within the longer masker files to be 1 s longer than the target sentence, allowing for a 500 ms onset and offset relative to the target sentence. Given that target sentences were around 2 s in duration on average, the combined stimuli were around 3 s in duration on average. The masker was mixed with the test sentence at the targeted SNR according to long-term RMS. A short call tone

before each stimulus alerted the listener. The resulting output signal was sent to a DA10 digital-to-analog converter of Lavry Engineering, Inc. (Washington, USA) and was diotically presented to the participant via Sennheiser HD600 headphones (Wedemark, Germany). After scoring the verbal response within the software, the next test sentence and masker excerpt were selected in the same way as above. The participants took a short break every three trials. In total, testing lasted approximately three hours, including screening, training, baseline measurements, data collection, and breaks. Testing was completed in one or two days, with no more than 3 weeks in between test sessions.

3. Results

Figure 2 shows boxplots for SRTs with the different maskers. In general, the SRTs were lower for young adults than older adults, indicating better performance. The largest mean masking difference between the two groups was for the speech masker (2.1 dB), followed by music maskers (0.4–1.6 dB), and steady noise (0.8 dB). SRTs also differed across maskers. A two-way mixed-design analysis of variance (ANOVA) was performed with participant group as between-subject factor (two levels: younger and older) and masker type as within-subject factor (six levels: male talker, Chopin, Trijntje, Cello, Borsato, and steady noise). Results showed significant main effects for group [$F(1,23) = 7.1$, $p = 0.014$, power = 0.722] and masker type [$F(5,115) = 239.2$, $p < 0.001$, power = 1.000], but no significant interaction [$F(5,115) = 2.6$, $p = 0.103$, power = 0.622). *Post hoc* pairwise comparisons (with Bonferroni correction) were performed to further examine the effect of masker. SRTs with Borsato were significantly poorer than with all other maskers ($p < 0.001$), then SRTs with steady noise were significantly poorer than with all the remaining maskers ($p < 0.001$). SRTs with Cello were significantly poorer than with the male talker ($p < 0.001$) and Chopin ($p < 0.001$). There were no significant differences in masking between the male talker, Chopin and Trijntje maskers ($p > 0.05$, after correction).

4. Discussion

Supporting the main hypothesis, the present data show that even when young and older listeners have a comparable speech perception in quiet (as confirmed by the baseline measurements), older listeners have greater difficulty understanding speech with a competing talker, background music, or noise, consistent with previous studies (Bergman *et al.*, 1976; Pichora-Fuller *et al.*, 1995; Tun *et al.*, 2002). However, it is striking that this age effect was observed even in middle-aged participants with no significant hearing impairment. A second hypothesis was that age effects would differ across maskers, with the speech and music maskers showing the largest age effect, and the steady noise the smallest age effect. While there was a trend that agreed with this hypothesis, there was no significant interaction between participant group and masker type.

Age effects were somewhat small, ranging from 0.4 to 2.1 dB differences in the mean SRT across maskers. However, given that the older participants were middle-

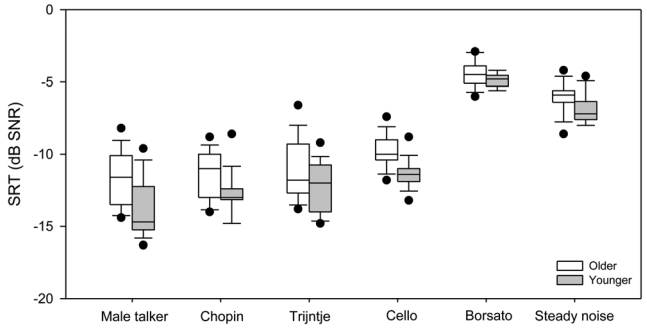

Fig. 2. Box plots for older (white boxes) and younger listeners (gray boxes). The *x* axis shows masker type, and the *y* axis shows the SRT. In each box, the solid line shows the median, the error bars show the 10th and 90th percentiles, and the black circles show outliers.

aged (51 to 63 years old), the present data suggest the beginning phases of age-related susceptibility to interference from background music (as well as speech and noise maskers). Because normal hearing status and audibility were reasonably controlled, participant age was expected to interact with energetic masking only minimally. The only masker that would cause only energetic masking (steady noise) produced the strongest masking, and the speech masker the weakest, as expected. The age effect was smaller with the steady noise, and larger with the speech masker, also as expected, however, this trend was not significant (due to lack of significant interaction group × masker). The present results somewhat differed than that of Rajan and Cainer (2008) who observed an age effect only with a slightly older group (7 adults, age range 59–69 years, mean 62 years) and only with multi-talker babble, but not a steady speech-shaped noise. Rajan and Cainer (2008) used stricter inclusion criteria for normal hearing, which resulted in a smaller number of participants than the present study but with better audiometric thresholds. Despite our control for normal hearing and audibility across listeners, the small difference in audiometric thresholds may have contributed to the difference in study outcomes. From the music maskers, the Borsato masker had a similar effect as the steady noise, perhaps due to its broader and flatter spectrum (Fig. 1), as well as the mixture of multiple instruments and lyrics (Ekström and Borg, 2011; Eskridge *et al.*, 2012; Gfeller *et al.*, 2012). Cello, with seven instruments and no lyrics, produced smaller masking but a larger age effect. Similarly, the slow-tempo maskers with single instruments, Chopin and Trijntje, produced the least amount of masking but large age effects, more similar to that with the male talker masker. As shown by the temporal envelopes in Fig. 1, the male talker, Chopin and Trijntje maskers all contained clear dips in amplitude, which may have provided listeners with glimpses of the target speech. Older participants seemed to take less advantage of such glimpsing, consistent with previous studies (Tun *et al.*, 2002; Helfer and Freyman, 2008; Saija *et al.*, 2014). Cognitive factors, such as working memory capacity, processing speed, inhibition of distractors, as well as divided and selective attention, likely play an important role in understanding speech in the presence of dynamic maskers (Pichora-Fuller *et al.*, 1995; Helfer and Freyman, 2008), and these may be already compromised by middle age (Park *et al.*, 2002; Humes *et al.*, 2006). The present data suggest that age effects do not result from energetic masking only, but instead from a mixture of energetic and informational masking, where cognitive factors may also play a role.

We have further analyzed our results to probe any potential confounding factors. It is unlikely that the masker and age effects were caused by a familiarity effect (Russo and Pichora-Fuller, 2008), as only familiar and generally popular music was selected as maskers. We have ruled out the bilingual exposure as a confounding factor, as the statistical results did not change after removing the few subjects who came from bilingual families (one from the younger group, three from the older group). A mixed-design ANOVA showed significant main effect of group [$F(1,19) = 10.0$, $p = 0.005$, power = 0.851], masker type [$F(5,95) = 217.6$, $p < 0.001$, power = 1.000], and different from the previous analysis, there was also an interaction between group and masker type [$F(5,95) = 2.5$, $p = 0.039$, power = 0.751]. Oxenham *et al.* (2003) had previously shown less informational masking in musicians compared to non-musicians, presumably due to better analytical listening skills. After removing the participants who played music (two from the younger group, five from the older group), a mixed-design ANOVA showed no significant effect of group [$F(1,16) = 3.6$, $p = 0.076$, power = 0.429], a significant effect of masker type [$F(5,80) = 161.2$, $p < 0.001$, power = 1.000], and no significant interaction [$F(5,80) = 0.7$, $p = 0.639$, power = 0.235]. Note that the insufficient power for the group effect makes the contribution of musicianship difficult to interpret. To better identify the musician effect in older populations, further research will be needed with more statistical power.

In summary, the present study shows that music maskers can be more distracting to older individuals, even if only middle-aged and without significant hearing impairment. Hence, public places should pay attention to the level of background music (Ekström and Borg, 2011; Gfeller *et al.*, 2012). For many places, this may not

jibe with the desired ambiance, but reducing volume only a small amount may produce less interference with speech understanding by middle-aged and older listeners.

Acknowledgments

We thank Qian-jie Fu and the Emily Fu Foundation for software support, and Etienne Gaudrain for feedback on earlier versions of this manuscript. This study was supported by a VIDI Grant No. 016.096.397 from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw), funds from Heinsius Houbolt Foundation, Rosalind Franklin Fellowship from University Medical Center Groningen, and is part of the Healthy Aging and Communication research program. J.J.G. was supported by Grant No. NIH R01-004993.

References and links

Başkent, D. (**2006**). "Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of the spectral channels," J. Acoust. Soc. Am. **120**, 2908–2925.
Benard, M. R., and Başkent, D. (**2013**). "Perceptual learning of interrupted speech," PLOS ONE **8**, e58149.
Bergman, M., Blumenfeld, V. G., Cascardo, D., Dash, B., Levitt, H., and Margulies, M. K. (**1976**). "Age-related decrement in hearing for speech: Sampling and longitudinal studies," J. Gerontol. **31**, 533–538.
Blood, A. J., and Zatorre, R. J. (**2001**). "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion," Proc. Natl. Acad. Sci. U.S.A. **98**, 11818–11823.
Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.
Durlach, N. I., Mason, C. R., Kidd, Jr., G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking (L)," J. Acoust. Soc. Am. **113**, 2984–2987.
Ekström, S., and Borg, E. (**2011**). "Hearing speech in music," Noise Health **13**, 277–285.
Eskridge, E. N., Galvin, J. J., III, Aronoff, J. M., Li, T., and Fu, Q.-J. (**2012**). "Speech perception with music maskers by cochlear implant users and normal hearing listeners," J. Speech Lang. Hear. Res. **55**, 800.
Fu, Q.-J., and Galvin, J. J., III (**2007**). "Perceptual learning and auditory training in cochlear implant recipients," Trends Amplif. **11**, 193–205.
Gfeller, K., Turner, C., Oleson, J., Kliethermes, S., and Driscoll, V. (**2012**). "Accuracy of cochlear implant recipients on speech reception in background music," Ann. Otol. Rhinol. Laryngol. **121**, 782–791.
Hanser, S. B., and Thompson, L. W. (**1994**). "Effects of a music therapy strategy on depressed older adults," J. Gerontol. **49**, 265–269.
Helfer, K. S., and Freyman, R. L. (**2008**). "Aging and speech-on-speech masking," Ear Hear. **29**, 87–98.
Humes, L. E., Lee, J. H., and Coughlin, M. P. (**2006**). "Auditory measures of selective and divided attention in young and older adults using single-talker competition," J. Acoust. Soc. Am. **120**, 2926–2937.
Milliman, R. E. (**1986**). "The influence of background music on the behavior of restaurant patrons," J. Consum. Res. **13**(2), 286–289.
Oxenham, A. J., Fligor, B. J., Mason, C. R., and Kidd, G., Jr. (**2003**). "Informational masking and musical training," J. Acoust. Soc. Am. **114**, 1543–1549.
Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A., and Smith, P. K. (**2002**). "Models of visuospatial and verbal memory across the adult life span," Psych. Aging **17**, 299–320.
Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise," J. Acoust. Soc. Am. **97**, 593–608.
Rajan, R., and Cainer, K. E. (**2008**). "Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational masking," Neuroscience **154**, 784–795.
Russo, F., and Pichora-Fuller, M. K. (**2008**). "Tune in or tune out: Age-related differences in listening to speech in music," Ear Hear. **29**, 746–760.
Saija, J. D., Akyürek, E. G., Andringa, T., and Başkent, D. (**2014**). "Perceptual restoration of degraded speech is preserved with advancing age," J. Assoc. Res. Otolaryngol. **15**, 139–148.
Shi, L.-F., and Law, Y. (**2010**). "Masking effects of speech and music: Does the masker's hierarchical structure matter?," Int. J. Audiol. **49**, 296–308.
Stephens, D. (**1996**). EU Work Group on Genetics of Hearing Impairment, European Commission Directorate, Biomedical and Health Research Programme Hereditary Deafness, Epidemiology and Clinical Research (HEAR). EU Work Group 1996, Infoletter 2.
Tun, P., O'Kane, G., and Wingfield, A. (**2002**). "Distraction by competing speech in young and older adult listeners," Psychol. Aging **17**, 453–467.
Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.