Frequency importance functions for audiovisual speech and complex noise backgrounds

Joshua G. W. Bernstein and Ken W. Grant

Army Audiology & Speech Center, Walter Reed Army Medical Center, Washington, DC, USA

Two studies investigated the dependence on listening condition of the relative importance of different regions of the frequency spectrum toward speech intelligibility. For consonant recognition, low-frequency speech information becomes more important under audiovisual (AV) than audio-alone (AA) conditions. The first study investigated whether this effect holds for broadband sentence materials using a correlation method designed to estimate frequency weighting functions for spectral profile analysis, but applied to speech. Preliminary results indicate a shift in the frequency-band importance function (FBIF) toward lower frequencies for AV sentences, consistent with the idea that the visual (V) signal provides place-of-articulation information complementary to the voicing and manner cues provided by the low-frequency auditory (A) channels. FBIFs for AA and AV speech may also change in multitalker noise where target-masker segregation is requisite to speech understanding. A second study tested the hypothesis that low frequencies should also be more important than high frequencies for avoiding informational masking (IM) because of the availability of strong pitch cues for segregation. Preliminary results support this hypothesis, showing a small but significant increase in IM with increasing frequency for bandpass-filtered speech. Overall, these results show that the frequency dependence of speech intelligibility depends on the type of background noise and whether V information is available. Systematically characterizing these effects may guide dynamic hearing-aid systems that shift the amplification spectrum for different listening situations.

INTRODUCTION

The main goal of a hearing aid is to benefit the speech perception abilities of impaired listeners. To this end, hearing-aid amplification algorithms often target those frequencies that are most important for speech intelligibility (e.g. Byrne *et al.*, 2001). While amplification algorithms attempt to improve speech intelligibility in quiet or stationary noise under AA conditions, listeners are often confronted with situations involving complex noise backgrounds and V speech cues. The FBIFs for speech may be very different under these common listening situations. Two hypotheses with respect to the FBIFs for speech are tested. First, under AV conditions, where the talkers face is visible, low frequencies may become more important relative to the AA case because the V signal and higher-frequency A channels provide redundant place-of-articulation cues (Grant and Walden, 1996). Second, with complex noise maskers, low frequencies may increase in importance due to the increasing importance of strong low-fre-

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.). ISBN: 87-990013-1-4. Print: Centertryk A/S.

quency pitch cues (Houtsma and Smurzynski, 1990) that facilitate concurrent source segregation (Darwin and Hukin, 2000). Furthermore, because V cues may provide cues to segregation in addition to phonetic information about the target speech itself, FBIFs for complex noise conditions may be affected by V cues.

EXPERIMENT 1: THE AV FREQUENCY-IMPORTANCE FUNCTION

Previous studies of consonant recognition have shown that low-frequency information becomes more important when lipreading cues are available, for narrow bands of speech presented in isolation (Grant and Walden, 1996) or simultaneously (Grant, 2005). This is thought to be due to the complementary nature of V and low-frequency audio speech cues: the V signal provides details about consonant place-of-articulation, while the low-frequency audio signal provides information about voicing and to some extent manner. However, a frequency dependence of the V benefit has not been observed for sentence materials using isolated frequency bands (Grant and Braida, 1991). It may be that the frequency dependence is more subtle for sentences than for isolated consonants, perhaps due to the presence of prosodic cues and multiple vowel contexts. FBIFs obtained for bandpass or narrow-band speech can differ from than those derived from broadband speech (Turner *et al.*, 1998). We hypothesized that the frequency dependence of the V benefit may be observed for the more natural case of broadband sentences.

The correlation method for estimating the FBIF (e.g. Turner *et al.*, 1998) assumes that the total speech information (I) is a weighted (W_i) sum of the information (B_i) available in each frequency band (i), corrupted by internal noise (ε): $I = \Sigma W_i B_i + \varepsilon$. The relative importance of each frequency region for speech is estimated by presenting speech tokens filtered into a number of different frequency bands. Each band is separately degraded by a randomly selected level of noise before being recombined and presented to the listener. On each trial, the listener's task is to identify the speech. A point biserial correlation is then computed between the listeners' response (correct versus incorrect) and the SNR in each band. The strength of each correlation represents the influence of each band on the listeners' responses. If speech intelligibility performance is proportional to the speech information present, then the correlation coefficients provide estimates of the individual W_i's.

The basic assumption underlying the use of this method for speech – that each frequency band provides independent information that is combined to form a percept associated with the broadband stimulus (Richards and Zhu, 1994) – is almost certainly invalid, as across-frequency interactions are known to contribute to speech understanding (e.g. Müsch and Buus, 2001). Furthermore, the wideband (WB) importance pattern is clearly dependent on how the frequency spectrum is partitioned (Calandruccio and Doherty, 2007), but the amount of speech information available in each separate filter band is difficult to measure due to the very poor intelligibility in noise yielded by each band presented in isolation. These considerations complicate the interpretation of each particular derived FBIF. However, the primary focus of this experiment was to determine the effect of V cues on the A FBIF. The influence of V information, deduced by *comparing* the AA and AV FBIFs, is less likely to depend on the particularities of the A stimulus which was identical in AA and AV conditions.

Methods

In a pilot test, three bandpass Butterworth filters with 48 dB/octave slopes were adjusted to yield approximately equal speech intelligibility for filtered IEEE (1969) sentences when each band was presented in isolation. The resulting 3-dB passbands of 100-685 (Band 1), 1250-1893 (band 2) and 3500-5825 Hz (Band 3) each yielded approximately 30% of keywords correctly identified in quiet. These filter cutoffs were separated by at least 0.85 octaves, thereby limiting band-on-band energetic masking when the bands were presented simultaneously in the experiment described below.

Each sentence was filtered into the three frequency bands. The level of the speech was set at 65 dB SPL before filtering, and speech bands 2 and 3 were each amplified by 10 dB after filtering to increase audibility. The three speech bands were presented simultaneously in a stationary Gaussian noise filtered into the same three bands. The correlation method requires adding noise at a signal to noise ratio (SNR) selected randomly and independently for each band on each trial. To reduce testing time and limit the number of sentences required for each listener, a new sampling of SNRs was made for each keyword in the sentence. The temporal locations of the five key words were marked by hand for each sentence, and three SNRs for each keyword (one for each frequency band) were selected at random from a uniform distribution with a 6-dB range (with a midpoint of -1 or -3 dB for each listener), yielding approximately 30 and 70% correct performance in the AA and AV conditions, respectively. Transitions between SNR levels for successive keywords were interpolated on a dB scale. Listeners verbally repeated back each sentence to the best of their ability, and the experimenter entered the number of correct keywords into the computer before the next sentence was presented. The audio signal processing was performed identically in the AA and AV conditions. The only difference between the two conditions was that the AV condition also presented a synchronized video of the talker to a 19-inch video screen one meter in front of the listener. Thirty-six 10-sentence blocks were presented (18 for each modality, in random order), for a total of 900 words each AA and AV. To date, five normal-hearing listeners have participated in the experiment.

Results

For each subject and condition, three point-biserial correlation coefficients were calculated between the S/N for each band and the outcomes (correct/incorrect) across the 900 words. The band-importance function was then determined by normalizing the three raw correlation coefficients to yield a sum of one. Thus, the relative importance of each band represents the contribution of each band's correlation to the total (Turner *et al.*, 1998). Each subject had two sets of correlations (AA and AV). The mean FBIFs across five normal-hearing listeners are plotted in Fig. 1. The AA results (closed circles) are dependent on how the spectrum was partitioned (Calandruccio and Doherty, 2007) and are therefore of limited value in describing a general intelligibility function for broadband speech. The aim of the current experiment was to compare

AA and AV importance functions under identical A stimulus parameters. The importance function for AV speech (open circles) was different from that of AA speech, with bands 1 and 3 more and less important, respectively. This observation was supported by a significant effect of band [$F(2,8) = 7.1$, $p<0.05$] on the difference between the AA and AV importance functions. Two-tailed paired t-tests comparing the AA and AV functions indicated a significant decrease in the importance of band 3 under AV presentation ($p<0.05$), while the increase in the importance of band 1 just failed to reach significance ($p = 0.056$).

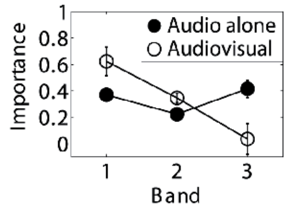

Fig. 1: Normalized correlations between SNR in each frequency band and performance. The importance function shifts toward low frequencies for AV speech.

Discussion

The effects of V cues on FBIFs are fairly consistent regardless whether consonants (Grant, 2005; Grant and Walden, 1996) or sentences (current study) are used as stimuli. However, the present experiment and the study by Grant and Braida (1991) are not in good agreement. Certain differences between the current experiment and that of Grant and Braida (1991) might explain why a frequency-dependent AV benefit was observed here and not in the previous study. First, the frequency dependence of speech intelligibility is generally more apparent under broadband (current study) than narrowband (Grant and Braida, 1991) listening conditions (Grant and Walden, 1996; Turner *et al.*, 1998; Calandruccio and Doherty, 2007). Across-frequency interactions in the broadband case may affect the measured AV benefit. Additionally, the steeper filter slopes employed by Grant and Braida (80 dB versus 48 dB/octave here) could have disrupted envelope modulations due to phase effects, leading to differences in AV integration.

The shift in the importance function toward low frequencies and away from higher frequencies under AV presentation supports the idea that the V signal provides the place-of-articulation information available in the high-frequency A band, such that the lower-frequency A information that is more poorly represented in the V signal tends to drive performance (Grant and Walden, 1996). This finding is also consistent with the observed increase in correlation as a function of frequency between A speech envelopes and V kinematic measurements (Grant and Seitz, 2000; Grant *et al.*, this volume), with high-frequency A channels providing envelope information that is mostly redundant with the V stimulus.

EXPERIMENT 2: FREQUENCY IMPORTANCE WITH COMPLEX MASKERS

Hearing-impaired (HI) and cochlear-implant (CI) listeners show particularly difficulty in competing-talker situations, where they do not receive the benefit relative to stationary noise enjoyed by NH listeners (e.g. Festen and Plomp, 1990). Because pitch information is thought to be important for simultaneous source segregation (e.g. Darwin and Hukin, 2000), it has been suggested that HI and CI listeners may benefit from highlighting the low frequencies (e.g. Chang *et al.*, 2006), where F0 is generally best represented (e.g. Houtsma and Smurzynski, 1990; Grant and Walden, 1996). However, Oxenham (this volume) found that NH listeners received similar benefit for a competing-talker masker relative to a stationary noise masker for both high- and low-pass filtered speech, shedding doubt on the idea that low-frequency information is essential for the benefit experienced by NH listeners.

One possible interpretation of this result is that many cues beside pitch exist for multitalker segregation (Darwin and Hukin, 2000) and that these cues are more uniformly distributed across the frequency spectrum. In situations of great perceptual similarity between masker and target (e.g. same-gender or same-talker interferers, where the masker and target sentence are produced by the same individual), cues to segregation are reduced. Such listening conditions can sometimes lead to performance deficits relative to the stationary noise case (Brungart, 2001), instead of the benefit (Festen and Plomp, 1990) sometimes seen in the interfering talker case. This deficit, thought to be central rather than peripheral in nature (Arbogast *et al.*, 2002), has been dubbed "informational masking" (IM). We hypothesized that the dependence on low-frequency pitch cues may become more important in the case of IM, where alternative cues to segregation are reduced.

To investigate this hypothesis, we investigated a situation with a large amount of IM, with target and masker speech generated by the same talker (Brungart, 2001). This was done for both AA and AV speech, with the AV condition providing measurable intelligibility at lower SNRs where IM is more likely to occur. For conditions with IM, the V cues available in many everyday listening situation can improve performance by as much as 9 dB (relative to 3.5 dB for the stationary noise case), while some IM still remains (Helfer and Freyman, 2005). Because V cues can provide some release from IM masking as well as information about the speech itself, the FBIF in complex noise may change under AV conditions.

Methods

Target speech stimuli consisted of the IEEE (1969) sentences spoken by a single female presented monaurally in noise, with uncorrelated contralateral masking noise presented to reduce acoustic and electric crosstalk. Four frequency conditions were tested: WB and three filtered conditions. In the filtered conditions, digital Butterworth filter order and cutoff frequencies were adjusted during pilot tests to yield roughly equal intelligibility performance across signal-to-noise ratio (SNRs) for each frequency band in the AA stationary noise condition (Band 1, 5th order lowpass, 0-1325 Hz; Band 2, 10th order bandpass, 1325-3050 Hz; band 3, 10th order bandpass, 3050-800 Hz).

Three different maskers were each spectrally shaped to have the same long-term power spectrum as the 720-sentence IEEE database. The first was a Gaussian stationary noise (N). A different random sample of noise was selected for each sentence presentation trial. A two-talker male (M) masker was generated by concatenating 24 sentences of the Hearing in Noise Test (HINT; Nilsson *et al.*, 1994). A two-talker same (S) female masker was generated by concatenating the first 120 IEEE sentences that formed the target speech set. The two-talker maskers were generated for each sentence presentation trial by summing two randomly-selected segments of the appropriate (M or S) long-duration masker. The target-masker onset delay (68-1041 ms, mean 620 ms, standard deviation 134 ms) was determined by the delay between the first video frame and the onset of the A signal for each recorded sentence. The masker always ended 250 ms following the target speech.

An adaptive paradigm measured speech intelligibility. Each sentence was repeated with SNR increasing in 3-dB steps until a threshold number of correct keywords was reached. A lower threshold was set for the filtered conditions (three out of five keywords) than the WB condition (four out of five) because some listeners in a pilot test sometimes had difficulty obtaining four correct words for the filtered stimuli. This paradigm gave a cue as to which sentence to attend to (especially in the S condition), with the target sentence (but not the interferers) remaining constant on successive presentations. Once threshold was reached or exceeded, the correct answer was displayed orthographically on the video screen, the SNR was reduced by 9, 12 or 15 dB (selected at random, with equal probability), and the process repeated with a new sentence. A run consisted of six sentences. If a listener achieved threshold on the first presentation of a given sentence, an additional sentence was added to the run. Seven NH listeners with normal or corrected-normal vision participated. Three listeners completed three runs, one completed two runs, and three completed one run per condition. Intelligibility was also measured in five NH listeners for the four frequency-band conditions presented in quiet; each listener achieved ≥ 94% correct for all four conditions.

In the filtered conditions, the WB target speech was set to 75 dB SPL, and the WB masker level adjusted to yield the desired nominal SNR. Target and masker were then both passed through the same filter. The target level for the WB condition was reduced to 65 dB SPL to limit the absolute masker level at unfavorable SNRs (-20 dB or less) sometimes encountered. The A stimuli were generated identically in AA and AV modalities; the only difference was that the AV conditions presented synchronized video of the target talker's face.

Results

A sigmoidal performance-intensity function was fit to the trial-by-trial data for each listener and condition. The functions were fixed at 100% correct for positive-infinite SNRs. For negative-infinite SNRs, the minimum value was fixed at 0% correct (AA) or allowed to vary as a single free parameter for each listener common to all AV functions. Small corrections ranging from -1.2 to +1.5 dB were applied to compensate for differences in spectral shape for the various noises in bands 2 and 3. Figure 2 plots estimated SRTs (SNR at 30% correct) as a function of frequency band, noise type and modality (AA or AV), averaged across the seven listeners. For all frequency-band conditions, the S masker generally yielded an SRT deficit relative to the N condition, indicating the presence of IM, while the M condition yielded an SRT benefit, consistent with previous results demonstrating a benefit from modulated maskers with little or no IM component (e.g. Festen and Plomp, 1990).

Estimates of the amount of IM in each condition are plotted in Fig. 3. Open and closed symbols represent the AA and AV conditions, respectively. The amount of IM can be estimated (Brungart, 2001) by comparing performance in the S condition to the M case, a modulated-noise masker condition with little or no expected IM. We consider the S-M comparison to be an upper-bound estimate of IM (upward-pointing triangles), due to the possibility of greater energetic masking in the S condition due to the similarity of the target and masker signals. By definition, the N conditions do not contain IM, but neither do they yield any modulated-masker benefit. Thus, the S-N comparison can be considered a lower-bound IM estimate (downward-pointing triangles). A significant main effect of band [$F(2,12) = 4.6$, $p<0.05$] for the combined upper- and lower-bound estimates confirmed the general trend for IM to increase with increasing frequency. This main effect held for the upper-bound estimates treated separately [$F(2,12) = 6.1$, $p<0.05$], but not for the lower-bound estimates ($p > 0.05$). There was also a significant effect of modality in the upper-bound case [$F(1,6) = 17.4$, $p<0.01$], reflecting an unexpectedly larger amount of IM estimated in the AV case.

Figure 4 shows the SRT difference between M and N conditions. Although visually

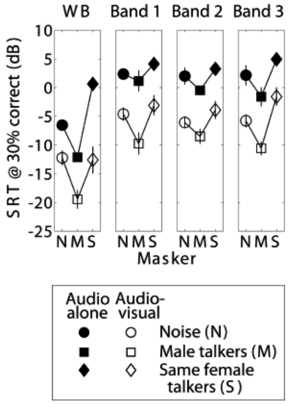

Fig. 2: Speech reception thresholds (SNR at 30% keywords correct) across frequency bands, noise types and modalities (AA or AV).

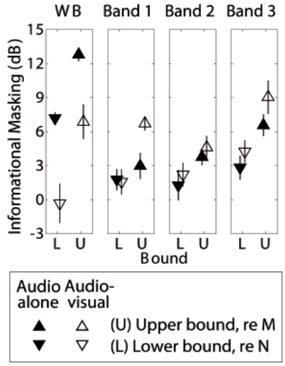

Fig. 3: Upper- and lower-bound IM estimates (see text for details). Band 3 shows more IM than the lower frequency bands.

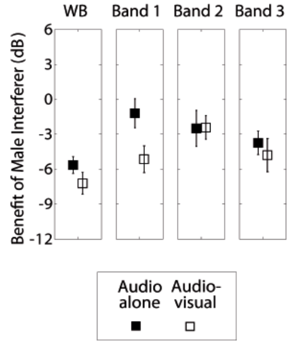

Fig. 4: SRT differences between M and N masker conditions. Negative values indicate a two-talker male benefit.

apparent in the data, the trend for the M benefit to increase with increasing frequency was not significant ($p=0.4$) in the AA condition, consistent with the results of Oxenham (this volume). However, in the AV condition, a main effect of band [$F(2,12) = 4.3$, $p<0.05$] reflects the greater amount of benefit seen in bands 1 and 3 relative to band 2.

Discussion

The data (Fig. 3) provide some subtle support for the hypothesis that IM should increase with increasing frequency due to the reduction in the strength of pitch cues. Nevertheless, the amount of IM may have been limited in both the AA and AV conditions, reducing our ability to observe a robust frequency dependence. In the AA conditions, IM may have been limited by the presence of level cues for SNRs greater than 0 dB. In the AV conditions, where SNRs were generally negative, level cues were less of an issue, which may explain how the unexpectedly greater amount of IM observed in the AV relative to the AA filtered conditions (Fig. 3). However, IM in the AV conditions was reduced by the V release from masking.

The reasons underlying the unexpected U-shaped benefit of the male talker (relative to the stationary noise condition) observed in the AV case remain speculative. One possibility is that modulation depth for speech may be greater at low and high frequencies than at mid frequencies (Greenberg *et al.*, 1998), yielding greater benefit for the modulated masker. A modulation analysis of the male-talker waveforms may shed light on this hypothesis. Alternatively, two different mechanisms may be responsible, benefiting the low and high frequencies, respectively. For example, temporal resolution tends to increase with increasing frequency (Snell *et al.*, 1994), potentially allowing listeners to make better use of the temporal masker fluctuations. The low-frequency benefit could arise under AV conditions if the information contained in the V signal about the target speech itself becomes more important in the modulated masker case (Helfer and Freyman, 2005), with V cues best complementing the low-frequency A information (Grant and Walden, 1996) as discussed in Section II. In either case, these effects may only have been observed in the AV conditions due to the larger effects of competing-talker maskers associated with lower SNRs (Oxenham, this volume).

CONCLUSIONS

Low frequencies are more important for speech intelligibility under AV than under AA conditions. Therefore, hearing aids and cochlear implants should focus on relaying the speech information available at these frequencies (e.g. voicing, nasality, and intonation) to help optimize speech intelligibility under AV conditions (Erber, 2003). This could be done by ensuring that these frequencies are present and fully audible, or by enhancing the information in these bands in some way (e.g. low-frequency envelope expansion, Apoux *et al.*, 2001). The frequency dependence of speech intelligibility for complex maskers is more complicated. Preliminary results indicate a subtle frequency effect, whereby high-frequency filtered speech shows more IM than low-frequency speech, consistent with the hypothesis that the stronger pitch cues at low frequencies provide IM release. In the case of interfering opposite-gender interferers where little or no IM is present, the frequency dependence of the benefit depended on the presence of V cues, favoring both high and low frequencies in the AV case, but showing no significant effect of frequency in the AA case. Taken together, these results indicate that the frequency dependence of speech intelligibility depends on the type of background noise and whether V cues are available. Hearing-aid signal processing algorithms may benefit if these effects are taken into account.

ACKNOWLEDGEMENTS

A grant from the Oticon Foundation supported this work. All participating listeners provided written informed consent prior to beginning the study. Preliminary results from experiment 1 were previously reported by Grant and Bernstein (2007). The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

REFERENCES

Apoux, F., Oliver, C., and Lorenzi, C. (**2001**). "Temporal envelope expansion of some speech in noise for normal-hearing and hearing-impaired listeners: Effects on identification performance and response times," Hear. Res. **153**, 123-131.

Arbogast, T. L., Mason, C. R., and Kidd, G. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086-2098.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101-1109.

Byrne, D., Dillon, H., Ching, T., Katsch, R., and Keidser, G. (**2001**). "NAL-NL1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures," J. Am. Acad. Audiol. **12**, 37-51.

Calandruccio, L., and Doherty, K. A. (**2007**). "Spectral weighting strategies for sentences measured by a correlation method," J. Acoust. Soc. Am. **121**, 3827-3836.

Chang, J. E., Bai, J. Y., and Zeng, F. G. (**2006**). "Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise," IEEE Trans. Biomed. Eng. **53**, 2598-2601.

Darwin, C. J., and Hukin, R. W. (**2000**). "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," J. Acoust. Soc. Am. **107**, 970-977.

Erber, N. P. (**2003**). "Use of hearing aids by older people: Influence of non-auditory factors (vision, dexterity)," Int. J. Audiol. 42, Suppl. **2**, 21-25.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725-1736.

Grant, K. W. (**2005**). "Frequency-band importance functions for auditory and auditory-visual speech recognition," J. Acoust. Soc. Am. **117**, 2424.

Grant, K.W. and Bernstein, J.G.W. (**2007**). "Frequency band-importance functions for auditory and auditory-visual sentence recognition," J. Acoust. Soc. Am., **121**, 3044 (Abstract).

Grant, K. W., Bernstein, J.G.W. and Grassi, E. (this volume). "Modeling speech intelligiblity," in Proceedings of the International Symposium on Auditory and Audiological Research (Helsingor, Denmark, August 29-31, 2007).

Grant, K. W., and Braida, L. D. (**1991**). "Evaluating the Articulation Index for audiovisual input," J. Acoust. Soc. Am. **89**, 2952-2960.

Grant, K. W., and Seitz, P. F. (**2000**). "The use of visible speech cues for improving auditory detection of spoken sentences," J. Acoust. Soc. Am. **108**, 1197-1208.

Grant, K. W., and Walden, B. E. (**1996**). "Evaluating the articulation index for auditory-visual consonant recognition," J. Acoust. Soc. Am. **100**, 2415-2424.

Greenberg, S., Arai, T., and Silipo, R. (**1998**). "Speech intelligibility derived from exceedingly sparse spectral information," in Proceedings of the International Conference of Spoken Language Processing (Sydney, Australia, December 1-4).

Helfer, K. S., and Freyman, R. L. (**2005**). "The role of visual speech cues in reducing energetic and informational masking," J. Acoust. Soc. Am. **117**, 842-849.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304-310.

IEEE (**1969**). IEEE recommended practice for speech quality measures (Institute of Electrical and Electronic Engineers, New York).

Müsch, H., and Buus, S. (**2001**). "Using statistical decision theory to predict speech intelligibility. I. Model structure," J. Acoust. Soc. Am. **109**, 2896-2909.

Nilsson, M., Soli, S., and Sullivan, J. A. (**1994**). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085-1099.

Oxenham, A.J. (this volume). "Pitch perception in normal, impaired and electric hearing," in Proceedings of the International Symposium on Auditory and Audiological Research (Helsingor, Denmark, August 29-31, 2007).

Richards, V. M., and Zhu, S. (**1994**). "Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients," J. Acoust. Soc. Am. **95**, 423-434.

Snell, K. B., Ison, J. R., and Frisina, D. R. (**1994**). "The effects of signal frequency and absolute bandwidth on gap detection in noise," J. Acoust. Soc. Am. **96**, 1458-1464.

Turner, C. W., Kwon, B. J., Tanaka, C., Knapp, J., Hubbartt, J. L., and Doherty, K. A. (**1998**). "Frequency-weighting functions for broadband speech as estimated by a correlational method," J. Acoust. Soc. Am. **104**, 1580-1585.