Role of temporal envelope and fine structure cues in speech perception: A review

CHRISTIAN LORENZI[1,2,3] AND BRIAN C. J. MOORE[4]

[1] *LPP CNRS, Université René Descartes Paris 5, Paris, France*

[2] *GRAEC, CNRS, France*

[3] *Dept d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France*

[4] *Department of Experimental Psychology, University of Cambridge, Downing street, Cambridge CB2 3EB, UK*

Over the last few decades, a variety of evidence has been presented to support the idea that, for normal-hearing listeners, both temporal envelope (E) and temporal fine structure (TFS) cues play a role in speech identification. E cues in a few frequency bands seem to be sufficient for good speech identification in quiet, but TFS cues appear to play an important role when background sounds are present, especially for "glimpsing" speech in the temporal minima of fluctuating background sounds. There is also evidence that cochlear damage associated with mild to moderate hearing loss may severely degrade the ability to use TFS cues while preserving the ability to use E cues in speech stimuli. This is consistent with the relatively preserved ability of hearing-impaired listeners to identify speech in quiet when audibility is controlled for, and the substantial deficits observed for these listeners when speech is masked by fluctuating background noise.

INTRODUCTION

All acoustic signals received by the auditory system are passed though an array of bandpass filters (the auditory filters). At the output of each filter, there are two forms of temporal information: fluctuations in the envelope (E, the relatively slow variations in amplitude over time) and fluctuations in the temporal fine structure (TFS, the rapid variations with rate close to the center frequency of the filter). This is illustrated for a speech sound in the companion paper by Moore (2008) in this volume. From a signal-processing point of view, TFS corresponds to the "carrier" signal while E corresponds to an amplitude modulator applied to the carrier. Over the last few decades, several psychophysical studies have led to the idea that there is an important dichotomy between E and TFS cues (e.g., Flanagan, 1980; Houtgast and Steeneken, 1985; Drullman, 1995; Smith *et al.*, 2002; Xu and Pfingst, 2003; Zeng *et al.*, 2004; Zeng *et al.*, 2005), especially in the context of pitch and speech perception.

Electrophysiological and brain-imaging studies conducted with humans and other mammals have revealed neurones in the brainstem and the auditory cortex that are sensitive to these two temporal features (e.g., Palmer, 1995; Shulze and Langner, 1997; Giraud *et al.*, 2000; Hart *et al.*, 2003; Joris *et al.*, 2004; Liegeois-Chauvel *et al.*,

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.

2004; Luo *et al.*, 2006). Psychoacoustical and neuropsychological studies conducted with speech and non-speech sounds have shown that cochlear hearing loss degrades the ability to encode and/or use TFS cues (e.g., Moore and Skrodzka, 2002; Buss *et al.*, 2004; Lacher-Fougère and Demany, 2005; Lorenzi *et al.*, 2006; Santurette and Dau, 2006) but preserves the ability to use E cues (e.g., Bacon and Viemeister, 1985; Turner *et al.*, 1995; Moore and Glasberg, 2001; Baskent, 2006), and that central damage (e.g., lesions of the primary and secondary auditory cortices) and some forms of language disorders (aphasia, amusia, and developmental dyslexia) are associated with a degradation in the ability to encode and/or use E cues (e.g., Griffiths *et al.*, 2000; Hescot et al, 2000; Lorenzi *et al.*, 2000a,b; Rocheron *et al.*, 2002; Füllgrabe *et al.*, 2004). Psychoacoustical studies of the role of TFS in normal and impaired hearing, especially in relation to "dip listening", are reviewed in a companion chapter in this volume (Moore, 2008).

PERCEPTION OF SPEECH STIMULI PROCESSED TO PRESERVE E OR TFS CUES

Normal-hearing listeners

Several researchers have investigated the role of E and TFS cues in speech identification tasks, using signal-processing techniques (so-called "vocoders") which preserve E cues while removing TFS cues and vice versa. Speech sounds were initially split into several contiguous frequency bands (also called analysis bands). E cues alone were preserved by extracting the envelope in each band and using the envelope to modulate the amplitude of a noise band or a sinusoid centered at the frequency of the band from which the envelope was derived. The modulated carriers were then combined (van Tasell *et al.*, 1987; Shannon *et al.*, 1995). Speech processed in this way will be referred to as "E-speech". These studies showed that, with a small number of bands (4-16), E cues can yield high levels of identification for speech presented in quiet (Drullman, 1995; Shannon *et al.*, 1995; Loizou *et al.*, 1999; Smith *et al.*, 2002). TFS cues alone were preserved by using the Hilbert transform (Hilbert, 1912) to extract the TFS in each band (e.g., Smith *et al.*, 2002). Essentially, the band signal was divided by the envelope magnitude at each instant in time. With this processing, each band signal becomes like a frequency-modulated (FM) sinusoidal carrier, with a constant amplitude. The band signals were then combined. Speech processed in this way will be referred to as "TFS-speech".

A problem with TFS speech was pointed out by Ghitza (2001). He showed that, although E cues are physically removed by the processing, they are reconstructed at the output of the peripheral auditory filters and may therefore contribute to intelligibility. This is especially the case when only a few relatively broad analysis bands are used. As a consequence, the identification of TFS-speech may be "contaminated" by a contribution of reconstructed E cues. This was demonstrated by the studies of Zeng *et al.* (2004) and Gilbert and Lorenzi (2006). They passed TFS-speech through a bank of simulated (gammachirp) auditory filters (Irino and Patterson, 2001). The reconstructed envelopes at the outputs of these filters were used in tone or noise vocoders, to assess the extent to which the reconstructed E cues supported speech recognition. The results of Gilbert and Lorenzi (2006) suggested that the reconstructed E cues did not play a major role in speech identification when the bandwidth of analysis filters was less than 4 ERBN (Glasberg and Moore, 1990), or equivalently, when the number of analysis bands was equal to or greater than 16 over the frequency range 0.08 to 8.02 kHz. In what follows, we focus on results obtained using a relatively large number of analysis bands, which were unlikely to be affected by reconstructed envelope cues.

When listeners were trained for a few hours, TFS-speech led to high levels of identification in quiet (Gilbert and Lorenzi, 2006; Lorenzi *et al.*, 2006; Gilbert *et al.*, 2007). Those high scores were relatively independent of level (i.e., 40 or 80 dBA) and were similar for TFS-speech that was unfiltered or highpass filtered at 0.4 kHz, suggesting that the important TFS cues for speech identification occur mainly for frequencies above 0.4 kHz (Sheft *et al.*, 2008).

Gilbert *et al.* (2007) studied the intelligibility of E- and TFS-speech processed using 16 bands. The processed stimuli were periodically interrupted at different rates. The periodic interruption was meant to act as a modulation masker, and was intended to interfere with the perceptual processing of the E cues below 16 Hz that are thought to be important for speech identification (e.g., Drullman *et al.*, 1994). The interrupted E- and TFS-speech stimuli were found to be highly intelligible in all conditions. However, when an effect of interruption rate was observed, the effect occurred at low interruption rates (< 8 Hz) and was stronger for E- than for TFS-speech, suggesting a greater involvement of modulation masking for E-speech. The different patterns of results obtained for E- and TFS-speech indicate that the two types of stimuli do not convey identical phonetic information, and confirm that identification of TFS-speech is not based on reconstructed E cues.

When a background sound such as a competing talker or a fluctuating noise is present, normally hearing listeners obtain much lower identification scores (than for intact speech) when TFS cues are removed by the use of a noise or tone vocoder (Nelson *et al.*, 2003; Qin and Oxenham, 2003; Stone and Moore, 2003; Zeng *et al.*, 2005; Füllgrabe *et al.*, 2006). This, combined with the results discussed above, indicates that the normal auditory system can use both E and TFS cues to achieve very good speech identification in quiet, but that TFS cues are required to segregate speech from fluctuating background sounds.

Hearing-impaired listeners

People with cochlear hearing loss typically have difficulty in understanding speech when background sounds are present, and the difficulty is especially marked when the background is fluctuating in some way (Duquesnoy, 1983). We present here evidence supporting the idea that the difficulty stems at least partly from a reduced ability to process TFS cues. One of the first studies investigating the relationship between TFS processing and speech perception was conducted by Buss *et al.* (2004). They assessed both FM detection and speech perception in quiet for listeners with mild to moderate cochlear hearing loss. FM detection thresholds were measured for an FM rate of

2 Hz and a sinusoidal carrier of 1 kHz, a condition where thresholds probably depend on the use of TFS cues, as coded by phase locking in the auditory nerve (Moore and Sek, 1996). Speech perception scores in quiet were measured for unfiltered speech and for speech lowpass filtered at 1.8 kHz. FM detection thresholds were higher than for normal-hearing listeners and were highly correlated ($r = -0.8$) with speech identification scores (for both unfiltered and filtered speech), consistent with the idea that sensorineural hearing loss may be associated with a reduced ability to use TFS information and that this may contribute to poor speech intelligibility.

Lorenzi *et al.* (2006) investigated this hypothesis more directly by measuring identification scores for unprocessed, E-, and TFS-speech in quiet for three groups of listeners: young with normal hearing and young and elderly with moderate “flat” hearing loss. After training, normal-hearing listeners scored perfectly with unprocessed speech, and about 90% correct with E- and TFS-speech. Both young and elderly listeners with hearing loss performed almost as well as normal with unprocessed and E-speech, but performed very poorly with TFS-speech, indicating a greatly reduced ability to use TFS. For the younger hearing-impaired group, scores for TFS-speech were highly correlated ($r = -0.8$) with the ability to take advantage of temporal dips in a background noise when identifying unprocessed speech. These results strongly support the idea that the ability to use TFS may be critical for “listening in the background dips”. Interestingly, a follow up study (Lorenzi *et al.*, 2008) demonstrated a similar pattern of results for E- and TFS-speech identification in quiet for listeners with high-frequency mild-to-severe hearing loss and normal (< 20 dB HL) audiometric thresholds below 2 kHz. In this study, E- and TFS-speech stimuli were lowpass filtered at 1.5 kHz in order to restrict their spectrum to the region where audiometric thresholds were normal. Only a few of the hearing-impaired listeners were able to score above chance (6.25%) for the TFS-speech, whereas normal-hearing listeners achieved scores of 20-50%. The results indicate that, for listeners with cochlear hearing loss, deficits in the ability to use TFS cues in speech can occur even when audiometric thresholds are within the normal range.

One problem with TFS-speech is that, during gaps in the speech in a particular analysis band, low-level noise in the recorded speech is amplified to the same level as the speech itself. This happens because the process of removing the E cues is equivalent to multi-channel compression with an infinite compression ratio; whatever the original envelope amplitude in a given band, the output envelope amplitude is constant. Bands with no speech information at a particular time are filled with distracting background sound. As a result, TFS-speech sounds harsh and very noisy. This may pose a particular problem to hearing-impaired listeners who, because of their broadened auditory filters, would suffer more from masking between channels. The problem becomes worse as the signal is split into more channels, as this would result in more masking across channels. Also, hearing-impaired listeners would be poorer at recovering any envelope cues that may still be available, again as a result of their broadened auditory filters. Together, these effects could account for some of the difference in performance between normal and hearing-impaired listeners.

Hopkins *et al.* (2008) adopted a different approach to assess the use of TFS information by normal-hearing and hearing-impaired listeners. Rather than creating signals that were intended to convey only TFS information, they measured performance as a function of the number of channels (analysis bands), N, containing TFS information; the other channels were noise or tone vocoded, so that they conveyed only E information. The rationale behind their approach is explained next.

Hopkins and Moore (2007) found that listeners with moderate cochlear hearing loss could make little use of TFS information to discriminate harmonic and frequency-shifted complex tones, when the tones contained only unresolved harmonics; for a summary of their work, see the chapter by Moore (this volume). If similar listeners were completely unable to use TFS information in speech, they would be expected to perform as well when listening to TFS-speech as when listening to unprocessed speech, provided that the number of processing channels was sufficiently large that the frequency selectivity of the processing was similar to or better than that of the peripheral auditory system of the listener. However, Baskent (2006) found that hearing-impaired listeners performed better in a phoneme-identification task when the syllables were unprocessed than when they were processed with a 32-channel noise-band vocoder. The disparity might arise because hearing-impaired listeners may be able to use TFS information at low carrier frequencies, but may be unable to use it at high frequencies. Hopkins and Moore (2007) showed that hearing-impaired listeners had a greatly reduced ability to discriminate the TFS of complex tones with unresolved components when all components were above 900 Hz, but they did not investigate sensitivity to TFS for lower frequencies. It is possible that listeners with moderate cochlear hearing loss have some ability to use TFS information below 900 Hz (Santurette and Dau, 2006), which could explain why they performed better in the unprocessed condition than in the 32-channel vocoded condition in the study of Baskent (2006). If listeners with moderate cochlear hearing loss can use TFS information only at low carrier frequencies, progressively replacing vocoded information with unprocessed information, starting at low frequencies, should improve performance only up to a cut-off frequency above which TFS information cannot be used. Hopkins *et al.* (2008) tested this hypothesis.

Speech reception thresholds (SRTs: the speech-to-background ratio required for 50% correct key words in sentences) were measured for signals that were unprocessed for channels up to and including channel number N and were vocoded for higher-frequency channels. The value of N was varied from 0 to 32 in steps of 4. A competing-talker background was used, because, as described earlier, TFS information may be particularly important for dip listening. Nine normal-hearing listeners and nine listeners with moderate cochlear hearing loss were tested. For the latter, the combined target and background signal were amplified using the "Cambridge formula" (Moore and Glasberg, 1998), which was intended to restore audibility as far as possible while avoiding excessive loudness. The frequency-gain characteristic was selected for each listener, based on the audiogram of the test ear. Listeners were trained for about one hour prior to formal testing. Figure 1 shows the mean data obtained using a noise vocoder, for both normal-hearing and hearing-impaired listeners. Mean SRTs are plotted as a function of N. The hearing-impaired listeners performed more poorly than the normal-hearing listeners in all conditions, but the difference in performance varied with N; for larger values of N, the difference in performance between the normal-hearing and hearing-impaired listeners was greater. The interaction between group membership and N was statistically significant. The improvement in performance going from completely vocoded stimuli (N = 0) to completely unprocessed stimuli (N = 32) was much greater for the normal-hearing listeners (14.5 dB) than for the hearing-impaired listeners (4.8 dB). For the latter, the improvement varied across listeners, with some listeners benefiting little, if at all, and others benefiting almost as much as the normal-hearing listeners. Hopkins *et al.* (2008) found a similar pattern of results when a tone vocoder rather than a noise vocoder was used.

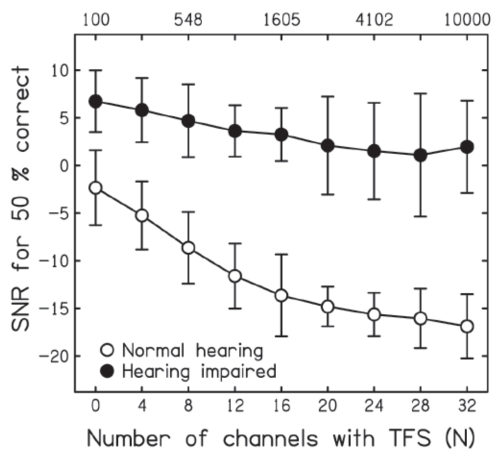

Fig. 1: Data from Hopkins *et al.* (2008) showing mean SRTs for normal-hearing and hearing-impaired listeners, plotted as a function of the number of channels with TFS information, N. The frequency corresponding to N is shown along the top axis in Hertz. Error bars show ± one standard deviation across listeners.

Reduced frequency selectivity has often been proposed as a factor contributing to the supra-threshold deficits associated with moderate cochlear hearing loss (Moore, 2007). Reduced frequency selectivity means that hearing-impaired listeners are more susceptible to masking across frequencies, and so partially explains why they perform poorly when listening in background sounds (Baer and Moore, 1993; 1994). The different patterns of performance for the normal-hearing and hearing-impaired listeners in the results of Hopkins *et al.* (2008) cannot be accounted for by differences in across-frequency masking. Similar amounts of masking would be expected in all of the conditions that were tested, so if deficits caused by cochlear hearing loss were only a result of across-frequency masking, a similar pattern of performance would have been expected for the normal-hearing and hearing-impaired listeners. Reduced spectral resolution may account for the differences in performance between the listener groups when N = 0, in which case no TFS information was available. However, the increasing difference between SRTs for the normally-hearing and hearing-impaired listeners as TFS information was added (as N was increased) is likely to reflect a different ability to use TFS information between the two groups.

In summary, the present results suggest that both E and TFS cues play an important role in speech identification. Either E or TFS cues appear sufficient (but not necessary) in quiet, although E cues seem to be more salient. For speech presented in background sounds, both E and TFS cues appear to be important. The results reviewed above also indicate that cochlear damage associated with mild to moderate hearing loss may severely disrupt the ability to use TFS cues while preserving the ability to use E cues in speech stimuli. A consequence of this is a normal or nearly normal ability to identify speech in quiet (provided that audibility is controlled for), but poor speech identification in fluctuating background noise.

ACKNOWLEDGEMENTS

The work of author BM was supported by the MRC, Deafness Research UK, and RNID.

REFERENCES

Bacon, S. P., and Viemeister, N. F. (**1985**). “Temporal modulation transfer functions in normal-hearing and hearing-impaired subjects,” Audiol., **24**, 117-134.

Baer, T., and Moore, B. C. J. (**1993**). “Effects of spectral smearing on the intelligibility of sentences in the presence of noise,” J. Acoust. Soc. Am. **94**, 1229-1241.

Baer, T., and Moore, B. C. J. (**1994**). “Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech,” J. Acoust. Soc. Am. **95**, 2277-2280.

Baskent, D. (**2006**). “Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels,” J. Acoust. Soc. Am. **120**, 2908-2925.

Buss, E., Hall, J. W., 3rd, and Grose, J. H. (**2004**). “Temporal fine structure cues to speech recognition and pure tone modulation in observers with sensorineural hearing loss,” Ear Hear., **25**, 242-250.

Drullman, R. (**1995**). “Temporal envelope and fine structure cues for speech intelligibility,” J. Acoust. Soc. Am. **97**, 585-592.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). “Effect of reducing slow temporal modulations on speech reception,” J. Acoust. Soc. Am. **95**, 2670-2680.

Duquesnoy, A. J. (**1983**). “Effect of a single interfering noise or speech source on the binaural sentence intelligibility of aged persons,” J. Acoust. Soc. Am. **74**, 739-743.

Flanagan, J. L. (**1980**). “Parametric coding of speech spectra,” J. Acoust. Soc. Am. **68**, 412-419.

Füllgrabe, C., Berthommier, F., and Lorenzi, C. (**2006**). “Masking release for consonant features in temporally fluctuating background noise,” Hear. Res. **211**, 74-84.

Füllgrabe, C., Maillet, D., Moroni, C., Belin, C., and Lorenzi, C. (**2004**). “Detection of 1st- and 2nd-order temporal envelope cues in a patient with left brain damage," NeuroCase **10**, 189-197.

Ghitza, O. (2001). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust. Soc. Am. **110**, 1628-1640.

Gilbert, G., and Lorenzi, C. (**2006**). "The ability of listeners to use recovered envelope cues from speech fine structure," J. Acoust. Soc. Am. **119**, 2438-2444.

Gilbert, G., Bergeras, I., Voillery, D., and Lorenzi, C. (**2007**). "Effects of periodic interruption on the intelligibility of speech based on temporal fine-structure or envelope cues," J. Acoust. Soc. Am. **122**, 1336-1339.

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., and Kleinschmidt, A. (**2000**). "Representation of the temporal envelope of sounds in the human brain," J. Neurophysiol. **84**, 1588-1598.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103-138.

Griffiths, T. D., Penhune, V., Peretz, I., Dean, J. L., Patterson, R. D., and Green, G. G. (**2000**). "Frontal processing and auditory perception," Neuroreport **7**, 919-922.

Hart, H. C., Palmer, A. R., and Hall, D. A. (**2003**). "Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex," Cereb. Cortex **13**, 773-781.

Hescot, F., Lorenzi, C., Debruille, X., and Camus, J. F. (**2000**). "Amplitude-modulation detection for broadband noise in a single listener with left-hemisphere damage," Brit. J. Audiol. **34**, 341-351.

Hilbert, D. (**1912**). Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen (Teubner, Leipzig).

Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055-1068.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. (submitted).

Houtgast, T., and Steeneken, H. J. M. (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069-77.

Irino, T., and Patterson, R. D. (**2001**). "A compressive gammachirp auditory filter for both physiological and psychophysical data," J. Acoust. Soc. Am. **109**, 2008-2022.

Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude modulated sounds," Physiol. Rev. **84**, 541-577.

Lacher-Fougère, S., and Demany, L. (**2005**). "Consequences of cochlear damage for the detection of interaural phase differences," J. Acoust. Soc. Am. **118**, 2519-2526.

Liégeois-Chauvel, C., Lorenzi, C., Trébuchon, A., Régis, J., and Chauvel, P. (2004). "Temporal envelope processing in the human left and right auditory cortices,"

Cereb. Cortex **14**, 731-740.
Loizou, P. C., Dorman, M., and Tu, Z. (**1999**). “On the number of channels needed to understand speech,” J. Acoust. Soc. Am. **106**, 2097-2103.
Lorenzi, C., Debruille, L., Garnier, S., Fleuriot, P., and Moore, B. C. J. (**2008**). “Abnormal auditory temporal processing for frequencies where absolute thresholds are normal,” J. Acoust. Soc. Am. (submitted).
Lorenzi, C., Dumont, A., and Füllgrabe, C. (**2000a**). “Use of temporal envelope cues by developmental dyslexics,” J. Speech Lang., Hear. Res. **43**, 1367-1379.
Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). “Speech perception problems of the hearing impaired reflect inability to use temporal fine structure,” Proc. Natl. Acad. Sci. USA **103**, 18866-18869.
Lorenzi, C., Wable, J., Moroni, C., Derobert, C., Frachet, B., and Belin, C. (**2000b**). “Auditory temporal envelope processing in a patient with left-hemisphere damage,” Neurocase **6**, 231-244.
Luo, H., Wang, Y., Poeppel, D., and Simon, J. Z. (**2006**). “Concurrent encoding of frequency and amplitude modulation in human auditory cortex: MEG evidence,” J. Neurophysiol. **96**, 2712-2723.
Moore, B. C. J. (**2007**). Cochlear Hearing Loss: Physiological, Psychological and Technical Issues (Wiley, Chichester).
Moore, B. C. J. (**2008**). “The role of temporal fine structure in normal and impaired hearing," in International Symposium on Auditory and Audiological Research,” edited by T. Dau, Holmens Trykkeri, Denmark.
Moore, B. C. J., and Glasberg, B. R. (**1998**). “Use of a loudness model for hearing aid fitting. I. Linear hearing aids,” Br. J. Audiol. **32**, 317-335.
Moore, B. C. J., and Glasberg, B. R. (**2001**). “Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners,” J. Acoust. Soc. Am. **110**, 1067-1073.
Moore, B. C. J., and Sek, A. (**1996**). “Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking,” J. Acoust. Soc. Am. **100**, 2320-2331.
Moore, B. C., and Skrodzka, E. (**2002**). “Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation,” J. Acoust. Soc. Am. **111**, 327–335.
Nelson, P. B., Jin, S.-H., Carney, A. E., and Nelson, D. A. (**2003**). “Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners,” J. Acoust. Soc. Am. **113**, 961-968.
Palmer, A. R. (**1995**). “Neural signal processing,” in Hearing, edited by B. C. J. Moore (Academic Press, San Diego).
Qin, M. K., and Oxenham, A. J. (**2003**). “Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers,” J. Acoust. Soc. Am. **114**, 446-454.
Rocheron, I., Lorenzi, C., Fullgrabe, C., and Dumont, A. (**2002**). “Temporal envelope perception in dyslexic children,” Neuroreport **13**, 1683-1687.
Santurette, S., and Dau, T. (**2006**). “Binaural pitch perception in normal-hearing and

hearing-impaired listeners," Hear. Res. **223**, 29-47.

Shannon, R., Zeng, F-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303-304.

Sheft, S., Ardoint, M., and Lorenzi, C. (**2008**). "Speech identification based on temporal fine structure: comparison of two speech coding schemes," J. Acoust. Soc. Am. (submitted).

Shulze, H., and Langner, G. (**1997**). "Periodicity coding in the primary auditory cortex of the Mongolian gerbil (Meriones unguiclatus): two different coding strategies for pitch and rhythm?," J. Comp. Physiol. **181**, 651-663.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature **416**, 87-90.

Stone, M. A., and Moore, B. C. J. (**2003**). "Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task," J. Acoust. Soc. Am. **114**, 1023-1034.

Tuner, C. W., Souza, P. E., and Forget, L.N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568-2576.

van Tasell, D., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). "Speech waveform envelope cues for consonant recognition," J. Acoust. Soc. Am. **82**, 1152-1161.

Xu, L., and Pfingst, B. E. (2003). "Relative importance of temporal envelope and fine structure in lexical-tone perception (L)," J. Acoust. Soc. Am. **114**, 3024-3027.

Zeng, F. G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y. Y., and Chen, H. (**2004**). "On the dichotomy in the auditory perception between temporal envelope and fine structure cues," J. Acoust. Soc. Am. **116**, 1351-1354.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. USA **102**, 2293-2298.