The importance of temporal fine structure for the intelligibility of speech in complex backgrounds

BRIAN C. J. MOORE

Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

Any complex sound that enters the normal ear is decomposed by the auditory filters into a series of relatively narrowband signals, each of which can be considered as a slowly varying envelope (E) superimposed on a more rapid temporal fine structure (TFS). It has been argued by several researchers that E information is sufficient for good intelligibility of speech in quiet, but that understanding speech in complex backgrounds, especially competing talkers, may require the use of TFS information. Hearing-impaired people appear to have a reduced ability to use TFS information, and this may partly account for their relatively poor ability to understand speech in complex backgrounds. Several studies are reviewed that assess these ideas, using both normal-hearing and hearing-impaired listeners. It is argued that TFS information may contribute to the perceptual segregation of speech from complex backgrounds, but that TFS information is probably not selectively involved in "dip listening". The ability to use TFS information appears to decline with increasing age from the twenties onwards, even when the audiogram remains normal, and this partly accounts for the reduced ability of older listeners to understand speech in complex backgrounds.

INTRODUCTION

When a complex broadband sound such as speech is analysed in the normal cochlea, the result is a series of narrowband signals, with a bandwidth about 12-15% of the centre frequency, each corresponding to one position on the basilar membrane. Each of these signals contains two forms of information: fluctuations in the envelope (E, the relatively slow variations in amplitude over time) and fluctuations in the temporal fine structure (TFS, the rapid oscillations with rate close to the centre frequency of the band). Information about the TFS is carried in the pattern of phase locking in the auditory nerve (spikes synchronised to a specific phase of individual cycles of the TFS), especially the inter-spike intervals. Information about the envelope is carried by changes in firing rate of the auditory nerve over time, and/or by phase locking to the envelope.

E information alone in a small number of frequency bands seems to be sufficient to give good intelligibility of speech in quiet (Shannon *et al.*, 1995). However, several authors have argued that TFS information in addition to E information may be important for understanding speech in complex backgrounds, especially

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.

backgrounds that contain temporal dips (Lorenzi and Moore, 2008; Moore, 2008b; Hopkins and Moore, 2009). It is well established that human listeners have the ability to listen in the dips of a background sound that is fluctuating in amplitude. This has been demonstrated using a range of signal-detection tasks; for reviews, see Moore (1988; 2008a). It has been proposed that this ability depends partly on the use of TFS information (Moore and Glasberg, 1987; Schooneveldt and Moore, 1987; Fantini and Moore, 1994; Moore, 2008a; Goldman *et al.*, 2010); changes in TFS during the dips in the background may provide a cue for the presence of the signal.

The ability of normal-hearing listeners to understand speech in a background sound is markedly better when the background is fluctuating in amplitude than when it is steady (Duquesnoy, 1983; Festen and Plomp, 1990; Hygge *et al.*, 1992; Peters *et al.*, 1998; Füllgrabe *et al.*, 2006), an effect that has been attributed to dip listening. For a fixed overall speech-to-background ratio (SBR), the difference in intelligibility between speech in a steady and a fluctuating background has been referred to as "masking release" (Füllgrabe *et al.*, 2006; Lorenzi *et al.*, 2006a). It has been proposed that dip listening for speech also involves the use of TFS information (Lorenzi *et al.*, 2006b; Hopkins and Moore, 2009), and that a reduced ability to process TFS contributes to the reduced masking release that is typically found for people with cochlear hearing loss (Lorenzi *et al.*, 2006b; 2006a; Hopkins *et al.*, 2008; Hopkins and Moore, 2010b; 2011) and users of cochlear implants (Nelson and Jin, 2002; 2004; Nelson *et al.*, 2003). However, this issue is somewhat controversial. Bernstein and Grant (2009) pointed out that the magnitude of masking release tends to decrease with increasing SBR. Masking release may be higher for normal-hearing people than for hearing-impaired people and people with cochlear implants because the first group are typically tested at lower SBRs than the last two groups. When hearing-impaired and normal-hearing listeners are tested at the same SBR, masking release is often similar for the two groups (Bernstein and Grant, 2009), although some studies still show a deficit in masking release for hearing-impaired and cochlear-implant listeners; for a review, see Léger *et al*. (2011, this volume).

The main issue addressed in this paper is whether TFS information plays a special role in the ability to extract speech information from the dips in a fluctuating background, or whether the ability to use TFS information is helpful in some other way, for example by improving overall performance or by helping in the perceptual segregation of the target speech from the background.

Some of the experiments described below involve processing of the signal so as to remove TFS information while preserving E information, for example, via vocoder processing. It should be noted that what this processing does is to remove the TFS information that was present in the *original signal*. When the processed signals are presented to a normal auditory system, their spectro-temporal properties are represented both as E and as TFS information in the auditory nerve. Thus, removal of the original TFS information does not mean that no TFS information is available in the auditory nerve. This is one reason (among several) why vocoder processing does not adequately simulate what is happening in an impaired auditory system or a cochlear implantee with little or no sensitivity to TFS. The main reason for using

vocoder processing is that the TFS evoked by the processed signal in a normal auditory system is different from the TFS evoked by the original signal; the latter is assumed to carry more information than the former.

IS TFS INFORMATION ESSENTIAL FOR DIP LISTENING?

To explore the importance of TFS information for the intelligibility of speech in a complex background, Hopkins *et al*. (2008) measured speech reception thresholds (SRTs) for a target talker in a background talker as a function of the frequency range over which TFS information was available. The signal was split into 32 1-ERB_N wide channels. Above a cut-off channel, *CO*, channels were tone or noise vocoded, to remove the original TFS information. Channels up to and including *CO* were not processed. Hopkins *et al*. found that, as *CO* was increased, SRTs decreased (improved) more for normal-hearing subjects than for subjects with cochlear hearing loss, suggesting that the former gain a substantial benefit from the original TFS information, while the latter were less able to use that information. However, for normal-hearing subjects, masking release does occur even for fully vocoded signals, suggesting that the TFS information from the original signal is not essential for dip listening, although it contributes to it (Hopkins and Moore, 2009).

Lunner *et al*. (2011) repeated the experiment of Hopkins *et al*. (2008) using different types of speech materials and a tone vocoder. In one experiment, the target speech was drawn from the Danish HINT sentences. These are similar to the materials used by Hopkins *et al*.; they have a somewhat unpredictable structure and are drawn from an open set. The results were similar to those of Hopkins *et al*.: when *CO* was increased from 0 (fully vocoded signal) to 32 (intact signal), the SRT for the young normal-hearing subjects decreased by about 7 dB (from –3 to –10.3 dB), while the SRT for the older hearing-impaired subjects decreased by only 3 dB (from 3.1 to –0.1 dB). However, when Lunner *et al*. used speech materials with a highly predictable structure drawn from a closed set (Dantale 2), the decrease in SRT with increasing *CO* was similar for the two groups, even though the hearing-impaired group performed more poorly overall. When *CO* was increased from 0 to 32, SRTs decreased from –14.9 to –18.0 dB for the normal-hearing group and from –3.7 to –7.7 dB for the hearing-impaired group.

Lunner *et al*. (2011) explained these results in the following way. Auditory spectrograms (Moore, 2003) show that speech has a sparse representation in the auditory system; the energy is high in only a few spectro-temporal regions, with low energy elsewhere (Darwin, 2009). When the speech of two talkers is mixed, there is often relatively little overlap between the spectro-temporal regions dominated by one talker and the regions dominated by the other talker. In this situation, the ability to identify the speech of the target talker is limited mainly by informational masking (Brungart *et al.*, 2001) rather than by energetic masking. The problem for the listener is to decide which spectro-temporal regions originated from one talker and which from the other. If the spectro-temporal regions emanating from one talker are artificially segregated using a priori knowledge, for example via an "ideal binary mask" (Brungart *et al.*, 2006; Wang *et al.*, 2009), intelligibility can be very high

even for adverse SBRs and even for hearing-impaired listeners. TFS information may be mainly useful for reducing informational masking, by providing cues that aid the perceptual segregation of the target and the background (Zeng *et al.*, 2005; Strelcyk and Dau, 2009). With speech material such as Dantale 2, perceptual segregation may be greatly aided by the closed-set material, and by the predictable grammatical and temporal structure of the sentences. In this case, the original TFS information in the signal may no longer be required or may be required to a lesser extent. This could account for the small TFS benefit found for the normal-hearing subjects for the Dantale 2 sentences. If the hearing-impaired subjects had reduced sensitivity to TFS, this could explain why these subjects obtained only a small benefit from TFS information, regardless of the type of speech material.

MANIPULATING TFS INFORMATION IN THE PEAKS AND DIPS

Stone *et al*. (2011) investigated the relative importance of TFS information presented in different parts of the dynamic range of IEEE sentences in a single-talker background. They filtered the stimuli into 30 1-ERB_N-wide frequency channels (Glasberg and Moore, 1990), resembling the channels that would be created in a normal cochlea, and then manipulated each channel signal so that it contained the intact channel signal in the peaks and E information only in the dips, or E information in the peaks and the intact channel signal in the dips. The intact channel signal contained both the original envelope, E_O, and the TFS, so it is denoted E_O+TFS. The extracted envelope was slightly modified from the original envelope and is denoted E´. The extracted envelope was applied to a tone carrier at the centre frequency of the channel, and the resulting signal is denoted $E´_{Carr}$. The relative level at which the TFS information was added (to the peaks or dips) was systematically varied by means of a "switching threshold". The switching was done using smooth transitions, via a cross-fade signal.

The processing is illustrated in Fig. 1. The bottom trace in each panel shows the cross-fade signal for the chosen value of the switching threshold. The top trace in each panel shows a channel signal with $E´_{Carr}$ in the peaks (dark gray segments) and E_O+TFS in the dips (light gray segments); this condition is denoted $E´_{Carr}/E_O$+TFS. The middle trace in each panel shows a channel signal with E_O+TFS in the peaks (light gray segments) and $E´_{Carr}$ in the dips (dark gray segments); this condition is denoted E_O+TFS/$E´_{Carr}$. When the switching threshold was −2 dB (top), a large portion of $E´_{Carr}$ was selected in the top trace (times from 0.062 to 0.0115 s). When the switching threshold was +3 dB (lower panel), the cross-fade signal spent more time at 1, and the transition occurred midway between 0.062 and 0.0115 s. It is hard from Fig. 1 to distinguish any difference between the TFS of $E´_{Carr}$ and that of E_O+TFS, but differences do occur and can be seen when the signals are analyzed on a finer time scale. After processing, all channel signals were added together.

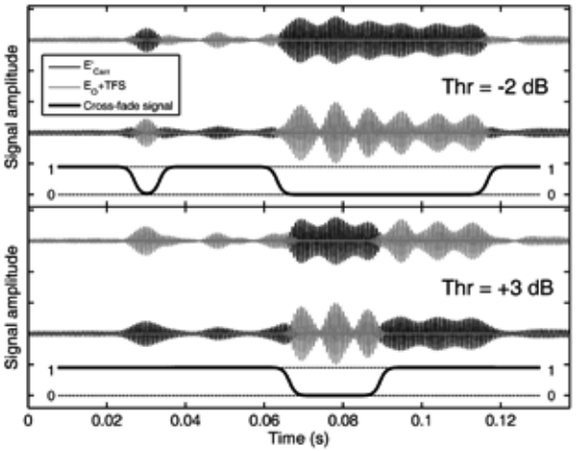

Fig. 1: Example of cross fading to produce the composite signal for channel 15 of a 30-channel system, for two values of the switching threshold (Thr): –2 dB (top) and +3 dB (bottom). The thick black line shows the cross-fade signal. In each panel, composite channel signals are shown for condition $E´_{Carr}/E_O$+TFS (upper trace) and condition E_O+TFS/$E´_{Carr}$ (lower trace). $E´_{Carr}$ segments are shown in dark gray and E_O+TFS segments in light gray.

In the experiment, the switching threshold was set to –99, –16, –9, –2, +5 and +99 dB relative to the root-mean-square channel level. Twenty sentences, comprising 100 keywords, were used to assess each condition. Fifteen young, normal-hearing subjects were tested in a counterbalanced order.

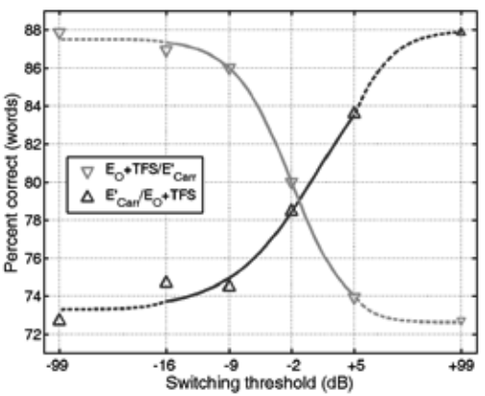

Fig. 2: Results of Stone *et al*. (2011) showing percent correct recognition of words in sentences for condition E_O+TFS/$E´_{Carr}$ (down-pointing triangles) and condition $E´_{Carr}/E_O$+TFS (up-pointing triangles), plotted as a function of switching threshold. The abscissa is distorted between –99 and –16 dB and between +5 and +99 dB.

The mean results across subjects for a SBR of 0 dB are shown in Fig. 2. Down- and up-pointing triangles denote the results for conditions E_O+TFS/E'_{Carr} and E'_{Carr}/E_O+TFS, respectively. The smooth curves are functions fitted to the data. For condition E_O+TFS/E'_{Carr}, performance worsened when the switching threshold increased above –16 dB and reached an asymptote when the switching threshold was between +5 and +99 dB. For condition E'_{Carr} /E_O+TFS, performance started to improve when the switching threshold was increased above –16 dB and continued improving when the switching threshold was above +5 dB.

An intensity-importance function (IIF) characterises the relative importance of information from different parts of the dynamic range of the signal (Studebaker and Sherbecoe, 2002; Bernstein and Grant, 2009; Stone *et al.*, 2010). The modulus of the slope of the smooth curves in Fig. 2, defined as the proportional change in intelligibility per dB of change in the switching threshold, gives IIFs for TFS information (assuming that the difference between E_O and E'_{Carr} does not have a material effect). IIFs for TFS are plotted in Fig. 3. The IIFs have been normalized so that their integral is unity. This makes the IIFs independent of the size of the change in performance between the extremes of the range of switching thresholds.

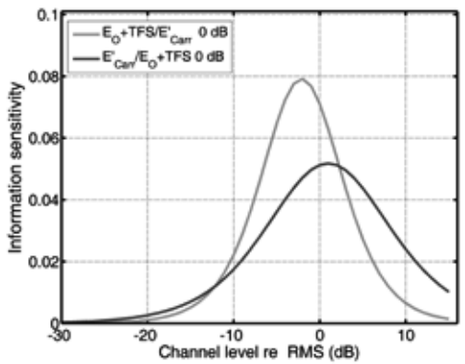

Fig. 3: IIFs for TFS derived from the fitted functions in Fig. 2. The light-gray curve is for condition E_O+TFS/E'_{Carr} and the dark-gray curve is for condition E'_{Carr}/E_O+TFS.

If TFS is especially important for extracting information from dips in the background, then the shapes of the IIFs for TFS information would be expected to reveal considerable importance for TFS in the lower-level portions of the dynamic range. The IIFs do not reveal such an effect. The IIFs show the greatest importance for relative levels between about –10 and +10 dB. The IIFs are similar to those derived for higher-rate E information by Stone *et al*. (2010). The maximum in the IIF for condition E_O+TFS/E'_{Carr}, for which TFS information was present in the peaks (light-gray curve), fell at a lower relative level than the maximum in the IIF for condition E'_{Carr}/E_O+TFS, for which E_O+TFS information was present in the dips (dark-gray curve), and the IIF for the former was somewhat sharper than that for the latter. Hence, for relative levels around –5 dB, the value of the IIF when TFS

information was present only in the peaks was markedly higher than when TFS information was present only in the dips. This means that TFS information at levels below the RMS level was used more effectively when TFS information was present in the peaks than when it was present in the dips. A possible explanation of this effect follows from the idea described earlier, that TFS information is mainly useful for perceptual segregation of the target from the background. TFS at relatively high levels may provide an initial basis for perceptual segregation, and this in turn may facilitate the use of TFS information at lower levels. When the addition of TFS information starts in the dips, the segregation process based on TFS may be initially more difficult; it does not "get off the ground" until TFS information is present over a reasonably wide range of levels.

THE EFFECT OF AGE ON THE ABILITY TO USE TFS INFORMATION

It was proposed many years ago that sensitivity to TFS reduces with increasing age (Pichora-Fuller and Schneider, 1991; 1992). Using two psychoacoustic tests of TFS sensitivity, the monaural TFS1 test (Moore and Sek, 2009) and the binaural TFS-LF test (Hopkins and Moore, 2010a), it has been shown that the decline occurs progressively with increasing age above about 20 years, even when audiometric thresholds are within the normal range (Hopkins and Moore, 2011; Moore *et al.*, 2011).

I describe next an unpublished study by Füllgrabe, Moore and Stone. They measured sensitivity to TFS using the TFS1 test at 1 and 2 kHz and the TFS-LF test at 0.5 and 0.75 kHz for a group of young (mean age = 23 yrs) and older (mean age = 67 yrs) subjects with matched verbal IQ. All audiograms were bilaterally normal (≤20 dB HL) for frequencies up to 6 kHz, and the mean audiograms were matched across age groups. Performance on the TFS tasks was significantly poorer for the older group than for the young group, consistent with the results described above. Note that auditory filters do not broaden with increasing age when the audiogram is normal (Peters and Moore, 1992), so the worse performance of the older group on the TFS tasks cannot be attributed to reduced frequency selectivity. Füllgrabe *et al*. also measured the intelligibility of consonants in steady noise and noise that was sinusoidally amplitude modulated at 5 or 80 Hz, using the same SBRs for the two groups. The scores for speech in noise (averaged across all noise types and all SBRs) were correlated with scores on the TFS tasks, averaged across all centre frequencies and tasks (expressed as discriminability index, d′, values). This is illustrated in the left panel of Fig. 4, which shows a scatter plot of consonant identification scores versus d′ scores for the TFS tasks: for the whole group, $r = 0.76$, $p < 0.01$; for the older subjects only, $r = 0.53$, $p < 0.05$; for the young subjects only, $r = 0.51$, ns.

The right panel of Fig. 4 shows mean scores for the identification of sentences in a single-talker background that was either co-located with the target speech or spatially separated from it (scores were averaged across the two cases). Again, there was a significant correlation between TFS scores and speech scores: for the whole group, $r = 0.83$, $p < 0.01$; for the older subjects only, $r = 0.59$, $p < 0.05$; for the young subjects only, $r = 0.87$, $p < 0.01$.

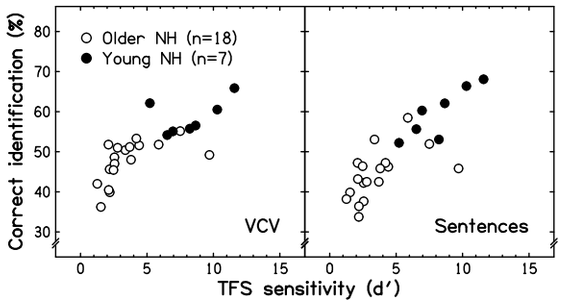

Fig. 4: Scatter plot of scores for consonants in vowel-consonant-vowel (VCV) nonsense syllables (left) and for sentences (right) against scores on the TFS tasks.

The results suggest that the poorer speech identification of the older subjects was associated with impaired TFS processing. However, the older subjects had poorer identification of speech in both the steady and modulated noises (Fig. 5). Thus, masking release was *not* reduced for the older subjects, despite their deficit in TFS processing.

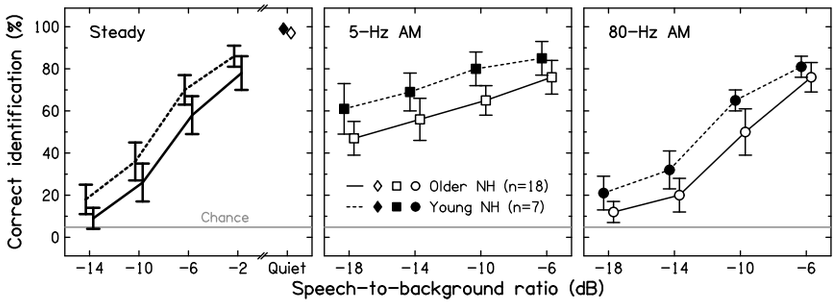

Fig. 5: Mean consonant-identification scores in quiet, and in steady, and 5- and 80-Hz 100% amplitude-modulated noise at different SBRs, for the two age groups. Error bars represent ±1 standard deviation about the mean.

If the ability to process TFS is specifically important for dip listening, then there should be a correlation between TFS scores and masking release. To assess whether this was the case, masking release for each subject was quantified as the difference in identification scores for consonants in steady noise, averaged across SBRs of –6, –10 and –14 dB, and in 5- and 80-Hz modulated noise, averaged across the same SBRs. Fig. 6 shows a scatter plot of masking release against TFS scores. There was no significant correlation between TFS scores and masking release: for the whole group, $r = -0.23$, ns; for the older subjects only, $r = -0.32$, ns; for the young subjects only, $r = -0.15$, ns. Overall, these results suggest, once again, that the ability to use TFS information affects the overall ability to understand speech in complex backgrounds but is not specifically involved in dip listening.

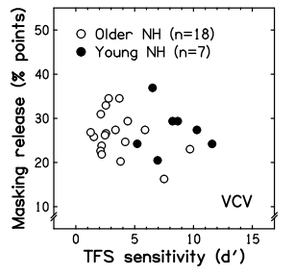

Fig. 6: Scatter plot of masking release against TFS scores.

CONCLUSIONS

The following findings suggest that the ability to use TFS information is not specifically involved in dip listening: (1) Masking release can occur when TFS information is severely degraded by vocoder processing; (2) Intensity-importance functions for TFS information do not indicate an undue importance of TFS in the lower parts of the speech dynamic range; (3) Older subjects with normal audiograms show a reduced ability to process TFS but show the same masking release as younger subjects when tested at the same SBRs; (4) Masking release for speech is not significantly correlated with a psychoacoustic measure of sensitivity to TFS.

These general conclusions are consistent with the results of a recent study by Bernstein and Brungart (2011). They showed that vocoder processing to remove the original TFS information for speech in various types of background sounds impaired overall performance. However, when the set size of the materials was adjusted to equate performance for the original and processed signals using the stationary noise masker, masking release was found to be the same for the original and processed signals. Thus, removal of the original TFS information does not reduce masking release.

TFS information from the original signal may contribute to speech perception by aiding the perceptual separation of the target and background. When the target speech has a highly predictable structure, this may aid the perceptual segregation of the target and background, reducing the importance of TFS information.

ACKNOWLEDGEMENTS

I thank Christian Füllgrabe, Kathryn Hopkins, Thomas Lunner, Aleksander Sek and Michael Stone for their collaboration in the work reported here and Christian Füllgrabe and Michael Stone for helpful comments. This research was supported by the MRC (UK), Deafness Research (UK) and the Oticon Foundation.

REFERENCES

Bernstein, J. G., and Brungart, D. S. (**2011**). "Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio," J. Acoust. Soc. Am. **130**, 473-488.

Bernstein, J. G. W., and Grant, K. W. (**2009**). "Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **125**, 3358-3372.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007-4018.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527-2538.

Darwin, C. J. (**2009**). "Listening to speech in the presence of other sounds," in *The Perception of Speech: From Sound to Meaning*, edited by B. C. J. Moore, L. K. Tyler, and W. D. Marslen-Wilsen (Oxford University Press, Oxford), pp. 151-169.

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.

Fantini, D. A., and Moore, B. C. J. (**1994**). "Profile analysis and comodulation detection differences using narrow bands of noise and their relation to comodulation masking release," J. Acoust. Soc. Am. **95**, 2180-2191.

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.

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

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

Goldman, S. A., Baer, T., and Moore, B. C. J. (**2010**). "Within-channel cues to comodulation masking release for single and symmetrically placed pairs of flanking bands," J. Acoust. Soc. Am. **128**, 2988-2997.

Hopkins, K., and Moore, B. C. J. (**2009**). "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise," J. Acoust. Soc. Am. **125**, 442-446.

Hopkins, K., and Moore, B. C. J. (**2010a**). "Development of a fast method for measuring sensitivity to temporal fine structure information at low frequencies," Int. J. Audiol. **49**, 940-946.

Hopkins, K., and Moore, B. C. J. (**2010b**). "The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **127**, 1595-1608.

Hopkins, K., and Moore, B. C. J. (**2011**). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," J. Acoust. Soc. Am. **130**, 334-349.

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. **123**, 1140-1153.

Hygge, S., Rönnberg, J., Larsby, B., and Arlinger, S. (**1992**). "Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds," J. Speech Hear. Res. **35**, 208-215.

Léger, A., Moore, B. C. J., and Lorenzi, C. (**2011**). "A review of speech masking release for hearing-impaired listeners with near-normal perception of speech in unmodulated noise maskers," in *Speech Perception and Auditory Disorders: 3rd International Symposium on Auditory and Audiological Research (ISAAR 2011)* (Centertryk A/S, Denmark), pp. (in press).

Lorenzi, C., and Moore, B. C. J. (**2008**). "Role of temporal envelope and fine structure cues in speech perception: A review," in *Auditory Signal Processing in Hearing-Impaired Listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007)*, edited by T. Dau, J. M. Buchholz, J. M. Harte, and T. U. Christiansen (Centertryk A/S, Denmark), pp. 263-272.

Lorenzi, C., Husson, M., Ardoint, M., and Debruille, X. (**2006a**). "Speech masking release in listeners with flat hearing loss: Effects of masker fluctuation rate on identification scores and phonetic feature reception," Int. J. Audiol. **45**, 487-495.

Lorenzi, C., Gilbert, G., Carn, C., Garnier, S., and Moore, B. C. J. (**2006b**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. USA **103**, 18866-18869.

Lunner, T., Hietkamp, R. K., Andersen, M. R., Hopkins, K., and Moore, B. C. J. (**2011**). "Effect of speech material on the benefit of temporal fine structure information in speech for normal-hearing and hearing-impaired participants," Ear Hear. (submitted).

Moore, B. C. J. (**1988**). "Dynamic aspects of auditory masking," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. Edelman, W. Gall, and W. Cowan (Wiley, New York), pp. 585-607.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing, 5th Ed.* (Emerald, Bingley, UK), pp. 413.

Moore, B. C. J. (**2008a**). "The role of temporal fine structure in normal and impaired hearing," in *Auditory Signal Processing in Hearing-Impaired Listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007)*, edited by T. Dau, J. M. Buchholz, J. M. Harte, and T. U. Christiansen (Centertryk A/S, Denmark), pp. 247-262.

Moore, B. C. J. (**2008b**). "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," J. Assoc. Res. Otolaryngol. **9**, 399-406.

Moore, B. C. J., and Glasberg, B. R. (**1987**). "Factors affecting thresholds for sinusoidal signals in narrow-band maskers with fluctuating envelopes," J. Acoust. Soc. Am. **82**, 69-79.

Moore, B. C. J., and Sek, A. (**2009**). "Development of a fast method for determining sensitivity to temporal fine structure," Int. J. Audiol. **48**, 161-171.

Moore, B. C. J., Vickers, D. A., and Mehta, A. (**2011**). "The effects of age on temporal fine structure sensitivity in monaural and binaural conditions," Int. J. Audiol. (submitted).

Nelson, P. B., and Jin, S. H. (**2002**). "Understanding speech in single-talker interference: Normal-hearing listeners and cochlear implant users," J. Acoust. Soc. Am. **111**, 2429.
Nelson, P. B., and Jin, S. H. (**2004**). "Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2286-2294.
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.
Peters, R. W., and Moore, B. C. J. (**1992**). "Auditory filters and aging: filters when auditory thresholds are normal," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 179-185.
Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577-587.
Pichora-Fuller, M. K., and Schneider, B. A. (**1991**). "Masking-level differences in the elderly: a comparison of antiphasic and time-delay dichotic conditions," J. Speech Hear. Res. **34**, 1410-1422.
Pichora-Fuller, M. K., and Schneider, B. A. (**1992**). "The effect of interaural delay of the masker on masking-level differences in young and old subjects," J. Acoust. Soc. Am. **91**, 2129-2135.
Schooneveldt, G. P., and Moore, B. C. J. (**1987**). "Comodulation masking release (CMR): effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. **82**, 1944-1956.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303-304.
Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (**2010**). "Relative contribution to speech intelligibility of different envelope modulation rates within the speech dynamic range," J. Acoust. Soc. Am. **128**, 2127-2137.
Stone, M. A., Moore, B. C. J., and Füllgrabe, C. (**2011**). "The dynamic range of useful temporal fine structure cues for speech in the presence of a competing talker," J. Acoust. Soc. Am. (in press).
Strelcyk, O., and Dau, T. (**2009**). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," J. Acoust. Soc. Am. **125**, 3328-3345.
Studebaker, G. A., and Sherbecoe, R. L. (**2002**). "Intensity-importance functions for bandlimited monosyllabic words," J. Acoust. Soc. Am. **111**, 1422-1436.
Wang, D., Kjems, U., Pedersen, M. S., Boldt, J. B., and Lunner, T. (**2009**). "Speech intelligibility in background noise with ideal binary time-frequency masking," J. Acoust. Soc. Am. **125**, 2336-2347.
Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. USA **102**, 2293-2298.

Controlling signal-to-noise ratio effects in the measurement of speech intelligibility in fluctuating maskers

JOSHUA G. W. BERNSTEIN

Audiology and Speech Center, Walter Reed National Military Medical Center, Bethesda, MD 20889, USA

The measurement of speech intelligibility in noise is often complicated by floor and ceiling effects. Because of this, adaptive methods are often used to determine the signal-to-noise ratio (SNR) required for a fixed performance level. Unfortunately, such methods relinquish control of the test SNR, confounding data interpretation when the effect of interest is SNR-dependent. For example, the intelligibility improvement afforded by glimpsing the target speech during brief dips in the level of a fluctuating masker is highly SNR-dependent. Thus, comparisons of performance in stationary and fluctuating maskers are susceptible to SNR confounds. Various methods of controlling SNR differences in the measurement of speech intelligibility are discussed, including the development and validation of a standardized intelligibility testing procedure that uses a variable response set size to control SNR differences. The application of these techniques to studies of hearing loss or simulated hearing loss demonstrate that impaired listeners may retain the ability to listen in the dips of a fluctuating masker to a much greater extent than previously thought.

INTRODUCTION

Normal-hearing (NH) listeners typically demonstrate better speech recognition when the target is presented in a fluctuating background (e.g., competing speech or modulated noise) then when it is presented at the same signal-to-noise ratio (SNR) in stationary noise (e.g., Miller and Licklider, 1950). This phenomenon, referred to as the fluctuating-masker benefit (FMB) or masking release, is thought to reflect dip listening – the ability to excise speech information during momentary dips in the masker level. Hearing-impaired (HI) listeners show little or no fluctuating-masker speech-reception advantage (e.g., Festen and Plomp, 1990), suggesting reduced dip-listening ability.

Signal processing methods have been used to simulate individual aspects of hearing loss in attempts to identify aspects of hearing loss that underlie the reduced FMB often observed for HI listeners. In particular, several studies have focused on the possible role of reduced frequency selectivity or an inability to use TFS information (i.e., fast timing information carried by phase locking in the auditory nerve) in limiting the FMB. These studies used spectral smearing algorithms to simulate reduced frequency selectivity (ter Keurs *et al*., 1993; Baer and Moore, 1994; Gnansia *et al*., 2009) or vocoding to remove TFS and present only the envelope (Qin and Oxenham, 2003; Gnansia *et al*., 2009; Hopkins and Moore, 2009). In each case,

Proceedings of ISAAR 2011: Speech perception and auditory disorders. 3rd International Symposium on Auditory and Audiological Research. August 2011, Nyborg, Denmark. Edited by T. Dau, M. L. Jepsen, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN 87-990013-3-0. EAN 9788799001330. The Danavox Jubilee Foundation, 2012.