Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal (L)

Christian Lorenzi[a)]
Equipe Audition (LPP CNRS, Universite Paris Descartes, and DEC, Ecole Normale Superieure), 29 rue d'Ulm, 75005 Paris, France and GRAEC, CNRS (Groupement de Recherche en Audiologie Experimentale et Clinique, GDR #2967), 29 rue d'Ulm, 75005 Paris, France

Louis Debruille, Stéphane Garnier, and Pierre Fleuriot
GRAEC, CNRS (Groupement de Recherche en Audiologie Expérimentale et Clinique, GDR #2967), 29 rue d'Ulm, 75005 Paris, France and Groupement d'Audioprothésistes Entendre, 65 Rue des 3 Fontanot, 92000 Nanterre, France

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 2 November 2007; revised 27 March 2008; accepted 6 May 2008)

The identification of nonsense syllables that were lowpass filtered at 1.5 kHz was compared for subjects with normal hearing and subjects with mild-to-severe hearing loss at high frequencies but with normal or near-normal hearing at low frequencies. Absolute thresholds were mostly within the normal range (<20 dB hearing level) for both groups for frequencies below 1.5 kHz. Performance was assessed with intact speech, speech that had been processed to preserve only temporal envelope cues in a few frequency bands (E speech), and speech that had been processed to remove envelope cues as far as possible while preserving temporal fine structure cues, again in a few frequency bands (TFS speech). For the intact speech and E speech, the hearing-impaired subjects performed slightly more poorly than the normal-hearing subjects, but this effect was significant only for the intact speech. For the TFS speech, the hearing-impaired subjects performed significantly more poorly than the normal-hearing subjects, with 12 out of 16 of the former performing at chance. The results indicate that, for people with hearing loss at medium to high frequencies, the processing of the TFS of speech can be degraded for frequencies where absolute thresholds are within the normal range.
 [DOI: 10.1121/1.2939125]

PACS number(s): 43.71.Ky, 43.71.Rt, 43.66.Sr [MW] Pages: 27–30

I. INTRODUCTION

Within the cochlea, broadband sounds like speech are decomposed by the "auditory filters" (Moore, 2003) into a series of narrowband signals, each evoked at a different place on the basilar membrane. Each signal can be considered as a "carrier"—the temporal fine structure (TFS), which is determined by the dominant frequencies in the signal that fall close to the center frequency of the band—and a temporal envelope (E), which corresponds to the relatively slow fluctuations in amplitude superimposed on the carrier. Information about speech sounds may be carried both by TFS and by E cues (Shannon *et al.*, 1995; Gilbert and Lorenzi, 2006; Hopkins *et al.*, 2008; Gilbert *et al.*, 2007).

Previously, we explored the role of TFS and E cues in speech perception for normal-hearing (NH) subjects and for young and older subjects with "flat" moderate hearing loss (Lorenzi *et al.*, 2006). Speech was filtered into 16 adjacent 0.35-oct-wide frequency bands spanning the range 0.08–8.02 kHz. The signal in each band was processed to preserve only E cues (by replacing the TFS with a sinewave at the center frequency of the band) or only TFS cues (by dividing the bandpass filtered signal by the envelope). The processed signals from the bands were then recombined. All groups scored 90–95% correct with E cues alone. After training, the NH group achieved about 90% correct with TFS cues alone, but both hearing-impaired (HI) groups performed very poorly, scoring between chance and 35% correct. The scores with TFS speech were highly correlated with the limited ability to take advantage of temporal dips when trying to understand unprocessed speech presented in a fluctuating background noise. We concluded that, regardless of age, hearing impairment adversely affects the ability to use TFS cues, and that such cues are important for the ability to listen in the dips of a competing background.

It is widely assumed that hearing impairment acts independently in different frequency regions, and that a "normal" audiometric threshold at a specific frequency, which is defined clinically as a threshold better than 20 dB hearing level (HL), implies normal auditory processing at that frequency. However, there is some evidence to the contrary. For example the ability to detect interaural time differences in low-frequency tones can be adversely affected by high-frequency cochlear hearing loss (Smoski and Trahiotis, 1986); for a review, see Moore (2007). Also, the ability to identify low-pass filtered speech can be adversely affected by high-frequency hearing loss (Horwitz *et al.*, 2002).

[a)]Author to whom correspondence should be addressed. Electronic mail: lorenzi@psycho.univ-paris5.fr

TABLE I. Age (years) and audiometric thresholds dB HL for each HI subject.

		Frequency (kHz)								
Listener	Age	0.25	0.5	0.75	1	1.5	2	3	4	6
HI1	67	20	25	30	30	30	35	55	55	55
HI2	53	15	10	15	10	15	20	35	50	35
HI3	46	30	30	35	30	35	35	45	60	60
HI4	39	20	15	20	25	25	30	45	40	85
HI5	58	30	35	35	35	40	40	35	35	30
HI6	54	25	30	30	30	35	35	40	50	75
HI7	38	20	15	15	15	25	50	55	50	50
HI8	61	15	15	15	20	25	20	30	30	55
HI9	63	15	15	10	10	20	25	45	55	85
HI10	61	15	15	15	20	20	20	50	55	55
HI11	60	15	15	20	20	25	30	40	55	35
HI12	68	10	5	10	15	15	35	65	70	70
HI13	59	15	10	15	20	20	45	55	70	100
HI14	57	15	15	20	15	15	35	50	50	65
HI15	55	5	5	10	10	20	30	35	45	55
HI16	56	15	15	10	15	15	45	55	60	70

Here we present further evidence that the processing of speech by HI listeners can be degraded when the speech is filtered so as to restrict its spectrum to the range where absolute thresholds are within normal limits. We also show that the degradation is greater for the processing of TFS cues than for the processing of E cues in speech.

II. METHOD

A. Subjects

Twelve subjects with normal hearing (NH1–NH12, age: 21–46 years, mean: 29 years) and 16 subjects with high-frequency cochlear hearing loss HI1–HI16 were tested. All NH subjects had audiometric thresholds between 0 and 20 dB HL at audiometric frequencies between 0.25 and 4 kHz at the test (right) ear. The ages of the HI subjects and the audiograms for the test (right) ears are shown in Table I. The HI subjects were chosen to have normal, or near-normal low-frequency hearing. However, the HI subjects had slightly poorer low-frequency hearing than the NH subjects. An analysis of variance (ANOVA) of the audiometric thresholds for frequencies below 2 kHz (0.25–1.5 kHz) with factors listener group (NH and HI), and audiometric frequency (0.25–1.5 kHz) showed significant main effects of group, $F(1,25)=26.734$, $p<0.0001$, and audiometric frequency, $F(4,100)=8.92$, $p<0.0001$, and a significant interaction between group and audiometric frequency, $F(4,100)=3.477$, $p<0.05$. The average threshold across all frequencies from 0.25 to 1.5 kHz (the PTA) was 6.8 dB HL for the NH group and 19.4 dB HL for the HI group. Despite this, 11 of the HI subjects (HI2 and HI7–HI16) had PTAs better than 20 dB HL, which is usually taken as the boundary between normal and impaired hearing.

B. Stimuli

Subjects were required to identify VCV nonsense syllables that were either intact, or were processed to preserve mainly envelope cues (E speech) or temporal fine structure cues (TFS speech). One set of 48 unprocessed VCV stimuli was recorded. There were three utterances of each of the 16 /aCa/ syllables (C=/p, t, k, b, d, g, f, s, ʃ, v, z, ʒ, m, n, r, l/) read by a French female speaker and digitized at a 44.1 kHz sampling frequency. Signals were digitized via a 16 bit analog-to-digital converter at a 44.1 kHz sampling frequency.

The processing was the same as described by Lorenzi *et al.* (2006) and by Gilbert and Lorenzi (2006). For all conditions (intact, E and TFS), each VCV signal was initially bandpass filtered using zero-phase, 3rd-order Butterworth filters into 16 adjacent, 0.35-octave-wide frequency bands spanning the range 0.08–8.02 kHz. This bandwidth corresponds to approximately 2 ERB_N (Glasberg and Moore, 1990) for midrange frequencies, but to about 1 ERB_N for low frequencies (Gilbert and Lorenzi, 2006). To create intact speech, all the band signals were summed. For the E and TFS conditions, the Hilbert transform was applied to the signal in each band to decompose the signal into its E (modulus of the analytic signal) and TFS (cosine of the argument of the analytic signal).

For the E condition, the envelope in each band was downsampled to 0.441 kHz, and then passed through a zero-phase, 6th-order Butterworth lowpass filter (cutoff frequency=64 Hz). The resulting envelopes were upsampled back to 44.1 kHz and used to amplitude modulate sinusoidal carriers with frequencies at the arithmetic center frequencies of the bands, and with random starting phase. The modulated signals were summed over the 16 frequency bands to form the E signal. For the TFS condition, the TFS for each band was multiplied by the root mean square (rms) power of the VCV in that band. This was done to avoid amplifying frequency bands carrying little or no speech information (e.g., bands dominated by recording noise). The "power-weighted" TFS signals were summed over the 16 frequency bands to form the TFS signal.

Following processing, the stimuli were lowpass filtered at 1.5 kHz (−72 dB/oct slope, Butterworth filter), so as to

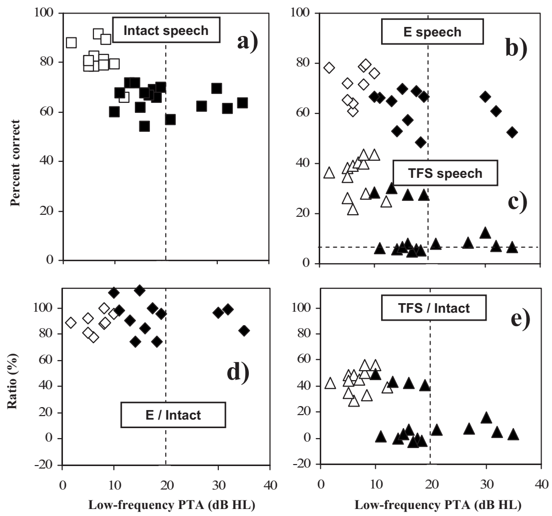

FIG. 1. Individual identification scores as a function of the PTA, for NH (open symbols), and HI subjects (filled symbols) for: (a) intact speech (squares); (b) E speech (diamonds); (c) TFS speech (triangles). For the ratio scores shown in panels (d) and (e), scores for intact, E and TFS speech were first transformed to RAU. The RAU-transformed scores for E and TFS speech were expressed relative to RAU-transformed scores for intact speech, and expressed as a percentage. Vertical dashed lines indicate the clinically accepted boundary between normal and impaired hearing.

restrict the spectrum to a frequency range where audiometric thresholds were <20 dB HL for 11 of the HI subjects and near normal for the rest. An advantage of this approach is that it eliminates possible confounding effects of differences in the widths of the auditory filters between groups; although auditory filters tend to broaden with increasing audiometric threshold, the broadening is typically small for hearing losses up to 30 dB (Moore, 2007). One problem with TFS speech is that the envelope in different frequency bands may be partly reconstructed in the auditory system (Gilbert and Lorenzi, 2006). The sharper the auditory filters, the greater is the degree of reconstruction. Given that the auditory filter widths were likely to be similar for the NH and HI subjects, it is very unlikely that any differences in sensitivity to TFS between the two groups could be attributed to differences in the degree of envelope reconstruction.

For the NH subjects, stimuli were presented at 75 dBA. For the HI subjects, a half-gain rule was applied, i.e., a gain equal to half of the average PTA was used. The resulting level was between 78 and 88 dBA, depending on the subject. The level was sufficient to ensure that all frequency components below 1.5 kHz were fully audible. Stimuli were presented using Sennheiser HD580 earphones. All subjects were trained for six, 5 min sessions, and performance was assessed over an additional four sessions. Subjects were tested first using intact and TFS speech, and were later recalled for the tests using E speech. Nine of the 12 NH subjects and 12 of the 16 HI subjects were able to return for the tests with E speech.

III. RESULTS

Figure 1 shows scores for the intact speech (a, squares) and the E and TFS speech (b, diamonds, and c, triangles, respectively), plotted as a function of the PTA. Scores for the NH and HI subjects are shown by open and filled symbols, respectively. A within-subjects ANOVA was conducted on the RAU-transformed scores with factors group (NH or HI) and processing (intact, E or TFS) (with missing values for the subjects who were not able to return for the tests with E speech). The main effect of group was significant: $F(1,19)=46.30$, $p<0.001$. The main effect of processing was also significant: $F(2,38)=272.5$, $p<0.001$. Finally, the interaction was significant: $F(2,38)=4.89$, $p=0.013$. *Post hoc* comparisons were based on the Tukey test.

For the intact speech [Fig. 1(a)], the NH subjects typically performed somewhat better than the HI subjects and the difference in mean scores was statistically significant ($p=0.02$). Similar moderate deficits in understanding lowpass filtered speech have been reported previously for HI subjects with good low-frequency hearing (Horwitz *et al.*, 2002). For the E speech [Fig. 1(b)], there was a slight tendency for the HI subjects to perform more poorly than the NH subjects, but the difference in mean scores was not statistically significant. This is consistent with previous work showing that HI subjects do not show a deficit in the processing of envelope cues in lowpass filtered speech (Turner *et al.*, 1995), although they do show a deficit for broadband speech processed using a large number of channels (Baskent, 2006), perhaps as a consequence of reduced frequency selectivity.

For the TFS speech [Fig. 1(c)], the NH subjects scored between 20 and 50%. These scores are much lower than those obtained when broadband TFS speech is used (Lorenzi *et al.*, 2006). This suggests that substantial TFS information can be used by the auditory system for frequencies above 1.5 kHz. Four of the HI subjects (HI2, HI11, HI13, and HI15) achieved scores of 25–30%, but most scored close to the chance level (6.25%), indicated by the horizontal dashed line. The difference in mean scores for the two groups was statistically significant: $p=0.0003$. These results indicate that many of the HI subjects had little or no ability to use TFS cues in lowpass filtered speech, despite their near-normal absolute thresholds.

Figure 1(d) shows, for each subject, the RAU-transformed scores for E speech expressed as a percentage of RAU-transformed scores for intact speech. These relative scores are somewhat more scattered for the HI than for the NH subjects, but most of the scores for both groups fall close to but a little below 100%. The means were 88% and 93% for the NH and HI groups, respectively. This indicates that the most salient cues for identification were the E cues, and removing TFS cues had only a small effect. Figure 1(e) shows, for each subject, the RAU-transformed scores for TFS speech expressed as a percentage of the RAU-

TABLE II. Mean identification scores across subjects (in percent correct, with standard deviation in parentheses) for each speech processing condition (Intact, E- and TFS-speech) and each filter slope/masker condition.

	Filter slope and masker	
Processing condition	−72 dB/oct (without masker)	−216 dB/oct (with masker)
Intact speech	85(4)	79(2)
E-speech	80(5)	68(5)
TFS-speech	46(9)	43(7)

transformed scores for intact speech. These relative scores for the NH subjects were much lower than the relative scores shown in panel d, confirming the greater salience of E cues. As expected, the relative scores for most of the HI subjects were very low, consistent with a greatly reduced ability to use TFS cues. The means were 44% and 14% for the NH and HI groups, respectively.

The scores of the HI subjects with TFS speech were not related to age ($r = 0.016$, not significant: NS). They were also not significantly related to the PTA ($r = -0.31$, NS), although it is noteworthy that the four subjects who achieved scores for TFS speech within the normal range all had PTA values below 20 dB. There were seven HI subjects with PTA values below 20 dB who scored close to the chance level with TFS speech. The four subjects who scored within the normal range for TFS speech did not differ from the other HI subjects in terms of their scores for intact speech or E speech.

IV. SUPPLEMENTARY EXPERIMENT: THE ROLE OF THE TRANSITION BAND

Although our stimuli were lowpass filtered at 1.5 kHz, subjects may have been able to use information from the "transition band" extending upwards from 1.5 kHz, where speech information was attenuated, but not removed completely. If the NH subjects could use this information more effectively than the HI subjects, this could account for the difference between the groups. This possibility was tested by running a new group of eight NH subjects (absolute thresholds below 20 dB HL at audiometric frequencies between 0.250 and 6 kHz) using conditions similar to those of experiment 1, and an additional condition in which the transition slope was steeper and a highpass-filtered noise was added to mask the transition band. The subject's ages ranged from 20 to 48 years.

The intact, E- and TFS-coded VCV stimuli were either generated exactly as for experiment 1 (with a lowpass filter slope of −72 dB/oct), or were lowpass filtered using a zero-phase, Butterworth lowpass filter with a cutoff frequency of 1.5 kHz and slope of −216 dB/oct and presented together with a speech-shaped noise masker that was highpass filtered (zero-phase, Butterworth filter with cutoff frequency of 1.5 kHz and slope of 108 dB/oct) at a signal-to-noise ratio of +12 dB (in terms of overall level). The subjects were initially trained with unprocessed, E and TFS broadband VCV stimuli, generated using a 16 band vocoder, until they achieved better than 70% correct for both E and TFS speech.

The procedure was identical to that for experiment 1. Mean results across listeners (based on 96 items per condition) are presented in Table II. A *t*-test with Bonferroni correction was used to estimate the significance of the differences in RAU scores across conditions. Preventing use of information from the transition band impaired speech identification for E-speech by 12 percent points ($p < 0.05$). In contrast, the effects were smaller for intact speech (6 percent points, NS) and TFS speech (3 percent points, NS).

Overall, these data suggest that NH subjects made little or no use of TFS information from the transition band. This indicates that the difference between the NH and HI subjects in their ability to use TFS information, as found in experiment 1, cannot be accounted for by the ability of the former to use information from the transition band.

V. CONCLUSIONS

Subjects with impaired hearing at high frequencies but normal or near-normal hearing at low frequencies often show a complete inability to use TFS information at low frequencies for speech reception. TFS information in the auditory nerve is probably not completely absent, since subjects with audiograms similar to those of the present subjects have some ability to discriminate changes in interaural phase of the TFS at the two ears for low-frequency amplitude-modulated tones (Lacher-Fougère and Demany, 2005). It may be that problems in using TFS occur when the TFS is complex and time varying, as in our TFS speech. In any case, the results demonstrate that normal audiometric thresholds at a specific frequency can be associated with strong abnormalities in the processing of TFS cues in speech at that frequency.

ACKNOWLEDGMENTS

The work of B.C.J.M. was supported by the MRC (UK). The authors thank Hedwig Gockel and two anonymous reviewers for helpful comments.

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.

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 interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues," J. Acoust. Soc. Am. **122**, 1336–1339.

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

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.

Horwitz, A. R., Dubno, J. R., and Ahlstrom, J. B. (**2002**). "Recognition of low-pass-filtered consonants in noise with normal and impaired high-frequency hearing," J. Acoust. Soc. Am. **111**, 409–416.

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.

Lorenzi, C., Gilbert, G., Carn, C., 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. U.S.A. **103**, 18866–18869.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).

Moore, B. C. J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*, 2nd ed. (Wiley, Chichester).

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

Smoski, W. J., and Trahiotis, C. (**1986**). "Discrimination of interaural temporal disparities by normal-hearing listeners and listeners with high-frequency sensorineural hearing loss," J. Acoust. Soc. Am. **79**, 1541–1547.

Turner, 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.