PNAS
Proceedings of the National Academy of Sciences of the United States of America www.pnas.org

Speech perception problems of the hearing impaired reflect inability to use temporal fine structure

Christian Lorenzi, Gaëtan Gilbert, Héloïse Carn, Stéphane Garnier, and Brian C. J. Moore

PNAS published online Nov 20, 2006;
doi:10.1073/pnas.0607364103

Speech perception problems of the hearing impaired reflect inability to use temporal fine structure

Christian Lorenzi*†‡, Gaëtan Gilbert*†, Héloïse Carn*†, Stéphane Garnier†§, and Brian C. J. Moore¶

*Equipe Audition, Laboratoire Psychologie de la Perception, Centre National de la Recherche Scientifique, Université René Descartes, and †Groupement de Recherche en Audiologie Expérimentale et Clinique Groupe de Recherche 2967, Departement d'Etudes Cognitives, Ecole Normale Supérieure, 29 Rue d'Ulm, 75005 Paris, France; §Groupement d'Audioprothésistes Entendre, 65 Rue des 3 Fontanot, 92000 Nanterre, France; and ¶Department of Experimental Psychology, Cambridge University, Cambridge CB2 3EB, United Kingdom

Edited by Michael M. Merzenich, University of California School of Medicine, San Francisco, CA, and approved October 12, 2006 (received for review August 24, 2006)

People with sensorineural hearing loss have difficulty understanding speech, especially when background sounds are present. A reduction in the ability to resolve the frequency components of complex sounds is one factor contributing to this difficulty. Here, we show that a reduced ability to process the temporal fine structure of sounds plays an important role. Speech sounds were processed by filtering them into 16 adjacent frequency bands. The signal in each band was processed by using the Hilbert transform so as to preserve either the envelope (E, the relatively slow variations in amplitude over time) or the temporal fine structure (TFS, the rapid oscillations with rate close to the center frequency of the band). The band signals were then recombined and the stimuli were presented to subjects for identification. After training, normal-hearing subjects scored perfectly with unprocessed speech, and were ≈90% correct with E and TFS speech. Both young and elderly subjects with moderate flat 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, TFS scores were highly correlated with the ability to take advantage of temporal dips in a background noise when identifying unprocessed speech. The results suggest that the ability to use TFS may be critical for ''listening in the background dips.'' TFS stimuli may be useful in evaluating impaired hearing and in guiding the design of hearing aids and cochlear implants.

background noise | dip listening | hearing impairment | speech intelligibility

When trying to understand speech in background sounds, hearing-impaired people perform somewhat more poorly than normal when listening in a steady background sound, but perform considerably more poorly when listening in a fluctuating background sound (1). Normal-hearing people get a considerable benefit from temporal dips in background sounds, but the hearing impaired seem to have little or no ability to "listen in the dips" to enhance speech perception, even when appropriate amplification is provided to ensure that the sounds in the dips are above the absolute threshold (2). This paper offers an explanation for why hearing-impaired people have such problems.

The auditory system performs a limited-resolution spectral analysis of sounds using an array of overlapping "auditory filters" with center frequencies spanning the range from ≈50 to 15,000 Hz (3, 4). The output of each filter is like a bandpass filtered version of the sound, which 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 center frequency of the band) (5). The TFS is often described as a "carrier" while the E is described as an amplitude modulator applied to the carrier. It is commonly believed that E cues are represented in the auditory system as fluctuations in the short-term rate of firing in auditory neurons, while TFS is represented by the synchronization of nerve spikes to a specific phase of the carrier (phase locking) (6, 7).

The role of E and TFS cues in speech perception has been studied by splitting sounds into contiguous frequency bands and processing the signal in each band so as to preserve only E or TFS cues. E cues alone can be presented by deriving the envelope in each band and using the envelope to amplitude modulate a noise band or sinusoid centered at the same frequency as the band from which the envelope was derived (8). With a moderate number of bands (4–16), E cues alone can lead to high intelligibility for speech in quiet both for normal-hearing people (9, 10) and for hearing-impaired people (11, 12). The fact that hearing-impaired people can achieve high levels of speech understanding for speech in quiet using E cues alone suggests that their difficulties in speech perception in noise do not stem from a basic difficulty in using E cues. However, for normal-hearing listeners, the intelligibility of speech based on E cues alone is very poor when a fluctuating background sound such as a single talker or an amplitude modulated noise is present (13–17); presumably, the same would apply for hearing-impaired listeners. This suggests that E cues alone do not allow effective listening in the dips of a background sound.

It is possible that the normal auditory system decides whether a speech signal in the dips of a background sound is produced by the target speech or is simply part of the background sound by using information derived from neural phase locking to TFS; changes in phase locking of auditory nerve discharges when a valley occurs indicate that a signal is present in the valley. This has been demonstrated mainly using basic psychoacoustic stimuli such as tones in fluctuating background sounds (18, 19). The difficulties experienced by hearing-impaired people when trying to listen in the dips may reflect a loss of ability to extract or use information from TFS, but, again, the loss of ability to use TFS has been demonstrated mainly using simple psychoacoustic stimuli (20–22) and the single study employing speech used an indirect approach based on correlation (23).

One way to study the role of TFS is to process speech so as to remove envelope cues as far as possible while preserving TFS cues, although this is technically difficult (16, 24). One approach is to filter the signal into a large number of contiguous bands, to extract the envelope in each band using the Hilbert transform (25), and to divide the signal in each band by the envelope (26). The resulting signal in each band has constant envelope amplitude, but a time-varying TFS. The band signals are then recom-

Author contributions: C.L., G.G., S.G., and B.C.J.M. designed research; G.G. and H.C. performed research; C.L. and H.C. analyzed data; and C.L., G.G., and B.C.J.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

‡To whom correspondence should be addressed at: Equipe Audition (LPP Centre National de la Recherche Scientifique–Paris 5 #2929), Department d'Etudes Cognitives, Ecole Normale Supérieure, 29 Rue d'Ulm, 75005 Paris, France. E-mail: lorenzi@psycho.univ-paris5.fr.

www.pnas.org/cgi/doi/10.1073/pnas.0607364103

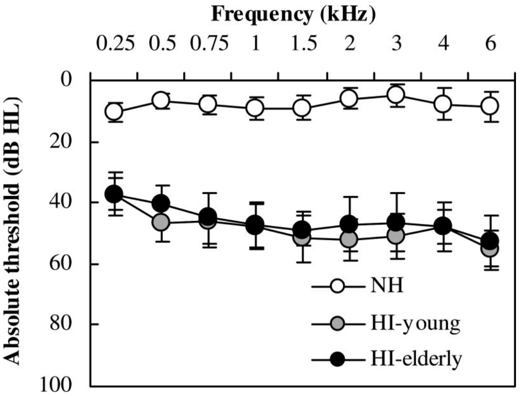

Fig. 1. Absolute thresholds (in dB HL) for the tested (right) ear as a function of frequency (in kHz). Mean data are presented for normal-hearing (open circles), and young (gray circles) and elderly hearing-impaired subjects (black circles). Error bars indicate ± 1 SD about the mean across subjects.

bined. A problem with this approach is that degraded E cues may be recovered at the outputs of the auditory filters because the E and TFS information are correlated (24, 27). However, this problem occurs mainly when the number of analysis bands is small, so that the widths of the analysis bands are larger than four times the widths of the normal auditory filters (28). When the number of bands is high, as used here, recovered E cues lead to identification of consonants in nonsense syllables that is close to chance for subjects with normal hearing (28).

We used a consonant-identification task to assess the ability of normal-hearing and young and elderly hearing-impaired subjects to identify intact speech and speech processed to preserve only E cues or TFS cues in 16 frequency bands (see Methods section). Mean audiograms for the three groups are shown in Fig. 1. Fig. 2*A* shows performance as a function of session number for TFS speech for normal-hearing subjects and young and elderly hearing-impaired subjects. Fig. 2*B* shows mean scores for all three groups for the last four test sessions. For the intact speech, the normal-hearing group achieved perfect scores and both groups of hearing-impaired subjects performed nearly as well as the normal-hearing group, although performance was slightly worse for the older group. For the E and TFS conditions, there was a significant interaction between factors processing scheme (two levels: E and TFS) and listener group (three levels); ($F_{2,18} = 80.9$; $P < 0.0001$). The normal-hearing group achieved scores over 90% for the E speech, after moderate training. Scores for the TFS speech also reached ≈90% correct, but these scores were found after more extensive training and were not significantly different from those for the E speech. These results confirm previous work suggesting that envelope cues are important for the intelligibility of speech in quiet (9, 29), but the results also indicate that TFS cues alone can give high intelligibility. The young hearing-impaired group performed slightly but not significantly better than the normal-hearing group for the E speech, but performed much more poorly than the normal-hearing group for the TFS speech (t (18) = 16.8; $P < 0.0001$). Most subjects scored close to the level that would be expected from the use of cues such as overall duration or long-term spectral cues (≈10–20% correct), and the two best-performing subjects achieved scores of ≈30% correct. The elderly hearing-impaired group performed slightly (but not significantly) less well than the other two groups for the E speech; but they performed very poorly for the TFS speech, with scores averaged over the last four sessions all below 20%. Mean scores for TFS speech were significantly lower for the elderly hearing-impaired group than for the normal-hearing group (t (18) = 18.62; $P < 0.0001$), but scores did not differ significantly for the young and old hearing-impaired subjects. Information transmission analysis (30) conducted on the confusion matrices indicated that reception of voicing, manner, and place of articulation in both groups of impaired subjects was barely degraded for E speech and nearly abolished for TFS speech.

These results indicate that moderate sensorineural hearing loss causes a dramatic deterioration in the ability to use TFS for speech perception. A modest deleterious effect of greater age was found but did not reach statistical significance, probably because of a floor effect. However, the ability to use envelope cues remains largely intact, explaining the preserved intelligibility of intact speech in quiet. As described in the introduction, E cues alone lead to poor intelligibility for normal-hearing subjects when a fluctuating background sound is present (13–17). This is consistent with the idea that TFS is important for listening in the dips and that the inability of hearing impaired subjects to use TFS cues is at least partly responsible for their reduced ability to understand speech in fluctuating backgrounds.

To assess whether the loss of ability to use TFS could explain the reduced ability of the hearing impaired to exploit dips in background sounds, the seven subjects from the young hearing-impaired group were assessed for their ability to understand intact nonsense syllables in a steady background noise and in a background noise that was sinusoidally amplitude modulated at an 8-Hz rate with a 100% depth, as described in (17). The speech-to-noise ratio was fixed individually at the level yielding

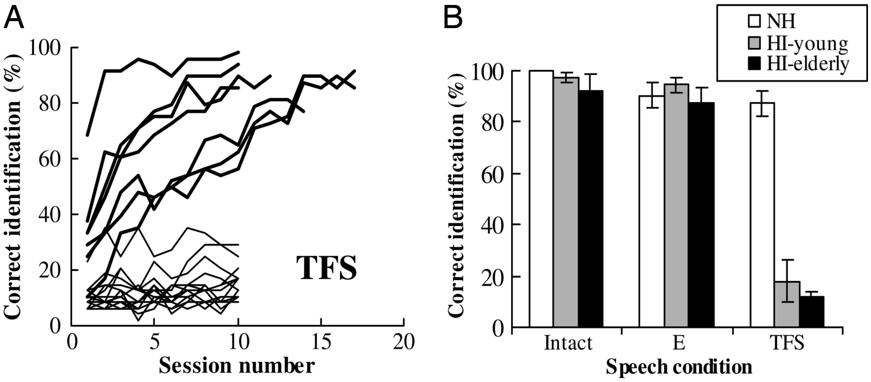

Fig. 2. Percent correct identification of types of speech. (*A*) Percent correct identification of TFS speech as a function of session number for normal-hearing subjects (thick lines), and young and elderly hearing-impaired subjects (thin lines). (*B*) Percent correct identification of intact, E and TFS speech for each group of subjects, averaged over the last four test sessions. NH, normal hearing; HI, hearing impaired. Error bars indicate ± one standard deviation about the mean across subjects.

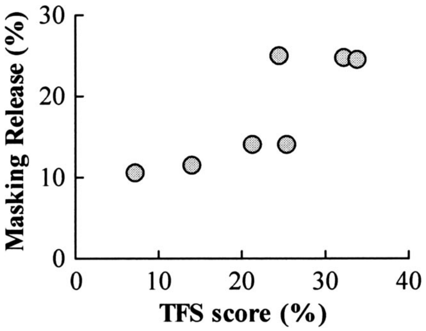

Fig. 3. Amount of masking release (in %) as a function of performance for TFS speech (in %) for the seven young hearing-impaired subjects.

≈50% correct identification for the speech in steady noise. The difference in scores for these two conditions provides a measure of the ability to listen in the dips of the modulated background; this is called masking release (17). TFS identification scores were assessed again in these subjects in four repeated sessions interleaved with the four sessions for the masking release task. Fig. 3 shows masking release plotted as a function of the mean score obtained for TFS speech. The correlation between the two is 0.83, which is significant at $P < 0.05$ (one-tailed test). Note that the amount of masking release for these subjects is markedly smaller than is typically found for normal-hearing subjects under similar conditions (17, 31). Overall, the pattern of the results supports the hypothesis that listening in the dips depends on the use of TFS information and that the greatly reduced benefit of dip-listening for hearing-impaired subjects is a result of the loss of ability to use TFS.

It is important to note that neural phase locking *per se* may not be affected by hearing impairment (32), although some studies do show alterations in phase locking to pure tones (33) and vowel-like sounds (34). However, the ability to use TFS cues may be impaired because the TFS information is extracted via cross-correlation of the outputs of different points along the basilar membrane (35–37). This cross-correlation process may be disrupted in hearing-impaired people because of abnormalities in the traveling wave on the basilar membrane caused by loss of the active mechanism (38). Another possibility is that the ability to use TFS cues is adversely affected by broadening of the auditory filters in hearing-impaired people (39). A consequence of such broadening is that the TFS within a specific channel can fluctuate more rapidly over time. It has been proposed that the mechanism that "decodes" TFS information is "sluggish" and cannot track rapidly fluctuating TFS (40, 41).

In summary, the results demonstrate a dramatic loss of the ability of hearing-impaired subjects to use TFS cues for speech perception. This loss appears to account for the fact that, unlike normal-hearing subjects, hearing-impaired subjects have little or no ability to exploit dips in background sounds to improve speech perception. The lack of ability to use TFS cues probably also limits the ability of people with cochlear implants to understand speech when background sounds are present. Improving the ability to use TFS should be a goal for designers of hearing aids and cochlear implants (42, 43), and TFS speech, as used here, may provide an important tool in evaluating how well this goal is achieved and in the early diagnosis of sensorineural hearing loss.

Methods

Stimuli. Speech signals were digitized (16-bit resolution) at a 44.1-kHz sampling frequency; they were then band-pass filtered using Butterworth filters (72 dB/oct rolloff) into 16 adjacent 0.35-oct wide frequency bands spanning the range 80–8,020 Hz. The bands were less than two times as wide as the "normal" auditory filters (44), and probably comparable to the widths of the auditory filters of the impaired subjects (45), thus ensuring that recovered E cues would be minimal (28) for both groups of subjects. The use of these analysis bands also ensured that the amount of spectral information provided by the E stimuli was similar for the normal-hearing and hearing-impaired subjects. The cutoff frequencies used and the technical details regarding stimulus generation are given in (28). These bandpass filtered signals were then processed in three ways. In the first (referred to as "intact"), the signals were summed over all frequency bands. These signals contained both TFS and E information. In the second (referred to as "E"), the envelope was extracted in each frequency band using the Hilbert transform followed by lowpass filtering with a Butterworth filter (cutoff frequency = 64 Hz, 72 dB/oct rolloff). The filtered envelope was used to amplitude modulate a sine wave with a frequency equal to the centre frequency of the band, and with random starting phase. The 16 amplitude-modulated sine waves were summed over all frequency bands. These stimuli contained only E information. In the third (referred to as "TFS"), the Hilbert transform was used to decompose the signal in each frequency band into its E and TFS components. The E component was discarded. The TFS in each band was multiplied by a constant equal to the root-mean-square (RMS) power of the bandpass filtered signal. The "power-weighted" TFS signals were then summed over all frequency bands. These stimuli contained TFS information only. In all conditions, the global RMS value of each stimulus was equalized.

All stimuli were delivered monaurally to the right ear via Sennheiser HD25–1 headphones. The stimuli were presented to the normal-hearing subjects at a level of 75 dBA and were presented to the hearing-impaired subjects after amplification of ≈20 dB to ensure that the stimuli were audible and comfortably loud.

Procedure. The identification task is fully described in (17, 28). A typical experimental session involved the identification of 48 vowel-consonant-vowel items (i.e., three exemplars of 16 /aCa/ utterances with C = /p, t, k, b, d, g, f, s, ʃ, v, z, j, m, n, r, l/, read by a French female speaker) processed using a given scheme and presented in random order in quiet. No feedback was given (even in the training sessions). For each subject and for each type of stimulus, training in sessions lasting 5 min was given until performance appeared to be stable (Fig. 2*A*). For a given processing scheme, stability was deemed to be achieved when identification scores covered a range of less than 9% across four successive sessions. All subjects were tested first using "intact" speech and thereafter using E and TFS speech. E and TFS conditions were interleaved until stability was reached in a given condition; testing was carried out after this in the remaining condition until stability was reached in the latter. More training was required for the TFS speech (10–17 sessions) than for the intact (4–6 sessions) or E speech (4–12 sessions).

Analyses. The percentage correct identification was calculated and a confusion matrix was constructed from the data for the last four sessions (192 responses) for a given processing scheme. All statistical analyses (ANOVA and *t*-tests with Bonferroni correction) were conducted on the arcsine-transformed scores obtained in the E and TFS conditions.

Subjects. Seven young subjects (mean age = 26; range: 21–35) with normal hearing (i.e., audiometric thresholds not exceeding 15 dB hearing level (HL) at any of the frequencies between 0.25 and 6 kHz) were tested (46). The hearing-impaired subjects were selected to have "flat" moderate hearing losses in the right ears, and they were divided into two groups: young (n = 7; mean age = 24; range: 18–37) and elderly (n = 7; mean age = 68; range:

63–72), because there is some evidence that the ability to use TFS decreases with increasing age (47, 48). The mean audiograms for all three groups are shown in Fig. 1. Air-conduction, bone-conduction, and impedance audiometry for the hearing-impaired subjects were consistent with sensorineural impairment. The origin of hearing loss was unknown for all elderly subjects and was either congenital or hereditary for the young ones. All impaired subjects had been fitted with a hearing aid on the tested ear for ≈9 years. All subjects were fully informed about the goal of the present study and provided written consent before their participation. This study was carried out in accordance with the French regulations governing biomedical research.

We thank Melodie Desaunay and Pierre Fleuriot for assistance in data collection and three anonymous reviewers for helpful comments on an earlier version of the manuscript. This work was supported by a grant from the Institut Universitaire de France (to C.L.) and by the Groupement de Recherche en Audiologie Expérimentale et Clinique (GDR Centre National de la Recherche Scientifique 2967). The work of B.C.J.M. was supported by the Medical Research Council of the United Kingdom.

1. Duquesnoy AJ (1983) *J Acoust Soc Am* 74:739–743.
2. Moore BCJ, Peters RW, Stone MA (1999) *J Acoust Soc Am* 105:400–411.
3. Fletcher H (1940) *Rev Mod Phys* 12:47–65.
4. Moore BCJ, Glasberg BR, Baer T (1997) *J Audio Eng Soc* 45:224–240.
5. Rosen S (1992) *Philos Trans R Soc London B* 336:367–373.
6. Rose JE, Brugge JF, Anderson DJ, Hind JE (1967) *J Neurophysiol* 30:769–793.
7. Joris PX, Yin TC (1992) *J Acoust Soc Am* 91:215–232.
8. Dudley H (1939) *J Acoust Soc Am* 11:169–177.
9. Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M (1995) *Science* 270:303–304.
10. Loizou PC, Dorman M, Tu Z (1999) *J Acoust Soc Am* 106:2097–2103.
11. Turner CW, Souza PE, Forget LN (1995) *J Acoust Soc Am* 97:2568–2576.
12. Souza PE, Boike KT (2006) *J Speech Lang Hear Res* 49:138–149.
13. Nelson PB, Jin SH, Carney AE, Nelson DA (2003) *J Acoust Soc Am* 113:961–968.
14. Qin MK, Oxenham AJ (2003) *J Acoust Soc Am* 114:446–454.
15. Stone MA, Moore BCJ (2003) *J Acoust Soc Am* 114:1023–1034.
16. Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargave A, Wei C, Cao K (2005) *Proc Natl Acad Sci USA* 102:2293–2298.
17. Füllgrabe C, Berthommier F, Lorenzi C (2006) *Hear Res* 211:74–84.
18. Schooneveldt GP, Moore BCJ (1987) *J Acoust Soc Am* 82:1944–1956.
19. Moore BCJ, Glasberg BR (1987) *J Acoust Soc Am* 82:69–79.
20. Moore BCJ, Moore GA (2003) *Hear Res* 182:153–163.
21. Lacher-Fougère S, Demany L (2005) *J Acoust Soc Am* 118:2519–2526.
22. Moore BCJ, Glasberg BR, Hopkins K (2006) *Hear Res*, in press.
23. Buss E, Hall JW, III, Grose JH (2004) *Ear Hear* 25:242–250.
24. Zeng FG, Nie K, Liu S, Stickney G, Del Rio E, Kong YY, Chen H (2004) *J Acoust Soc Am* 116:1351–1354.
25. Hilbert D (1912) *Grundzüge einer Allgemeinen Theorie der linearen Integralgleichungen* (Teubner, Leipzig).
26. Smith ZM, Delgutte B, Oxenham AJ (2002) *Nature* 416:87–90.
27. Ghitza O (2001) *J Acoust Soc Am* 110:1628–1640.
28. Gilbert G, Lorenzi C (2006) *J Acoust Soc Am* 119:2438–2444.
29. Drullman R (1995) *J Acoust Soc Am* 97:585–592.
30. Miller GA, Nicely PE (1955) *J Acoust Soc Am* 27:338–352.
31. Gustafsson HÅ, Arlinger SD (1994) *J Acoust Soc Am* 95:518–529.
32. Harrison RV, Evans EF (1979) *Arch Otolaryngol* 224:71–78.
33. Woolf NK, Ryan AF, Bone RC (1981) *Hear Res* 4:335–346.
34. Miller RL, Schilling JR, Franck KR, Young ED (1997) *J Acoust Soc Am* 101:3602–3616.
35. Loeb GE, White MW, Merzenich MM (1983) *Biol Cybern* 47:149–163.
36. Shamma SA (1985) *J Acoust Soc Am* 78:1622–1632.
37. Carney LH, Heinz MG, Evilsizer ME, Gilkey RH, Colburn HS (2002) *Acta Acustica Acustica* 88:334–346.
38. Robles L, Ruggero MA (2001) *Physiol Rev* 81:1305–1352.
39. Glasberg BR, Moore BCJ (1986) *J Acoust Soc Am* 79:1020–1033.
40. Sek A, Moore BCJ (1995) *J Acoust Soc Am* 97:2479–2486.
41. Moore BCJ, Sek A (1996) *J Acoust Soc Am* 100:2320–2331.
42. Nie K, Stickney G, Zeng FG (2005) *IEEE Trans Biomed Eng* 52:64–73.
43. Throckmorton CS, Selin Kucukoglu M, Remus JJ, Collins LM (2006) *Hear Res* 218:30–42.
44. Glasberg BR, Moore BCJ (1990) *Hear Res* 47:103–138.
45. Moore BCJ, Glasberg BR (2004) *Hear Res* 188:70–88.
46. ISO 389-8 (2004) *Acoustics—Reference Zero for the Calibration of Audiometric Equipment* (International Organization for Standardization, Geneva), Part 8.
47. Gordon-Salant S, Fitzgibbons PJ (1993) *J Speech Hear Res* 36:1276–1285.
48. Strouse A, Ashmead DH, Ohde RN, Grantham DW (1998) *J Acoust Soc Am* 104:2385–2399.