JARO
Journal of the Association for Research in Otolaryngology

Research Article

Temporal-Envelope Reconstruction for Hearing-Impaired Listeners

Christian Lorenzi[1], Nicolas Wallaert[1], Dan Gnansia[2], Agnès Claire Leger[1], David Timothy Ives[1], André Chays[3], Stéphane Garnier[4], and Yves Cazals[5]

[1]*Equipe Audition (CNRS, Universite Paris Descartes, Ecole normale superieure), Institut d'Etude de la Cognition, Ecole normale superieure, Paris Sciences et Lettres, 29 rue d'Ulm, 75005 Paris, France*

[2]*Neurelec, 2720 Chemin Saint Bernard, 06224 Vallauris Cedex, France*

[3]*Hôpital Robert Debré, Reims, France*

[4]*Groupement Entendre SAS, Paris, France*

[5]*Université Paul Cézanne, UMR CNRS 6231, 13916 Marseille, France*

ABSTRACT

Recent studies suggest that normal-hearing listeners maintain robust speech intelligibility despite severe degradations of amplitude-modulation (AM) cues, by using temporal-envelope information recovered from broadband frequency-modulation (FM) speech cues at the output of cochlear filters. This study aimed to assess whether cochlear damage affects this capacity to reconstruct temporal-envelope information from FM. This was achieved by measuring the ability of 40 normal-hearing listeners and 41 listeners with mild-to-moderate hearing loss to identify syllables processed to degrade AM cues while leaving FM cues intact within three broad frequency bands spanning the range 65–3,645 Hz. Stimuli were presented at 65 dB SPL for both normal-hearing listeners and hearing-impaired listeners. They were presented as such or amplified using a modified half-gain rule for hearing-impaired listeners. Hearing-impaired listeners showed significantly poorer identification scores than normal-hearing listeners at both presentation levels. However, the deficit shown by hearing-impaired listeners for amplified stimuli was relatively modest. Overall, hearing-impaired data and the results of a simulation study were consistent with a poorer-than-normal ability to reconstruct temporal-envelope information resulting from a broadening of cochlear filters by a factor ranging from 2 to 4. These results suggest that mild-to-moderate cochlear hearing loss has only a modest detrimental effect on peripheral, temporal-envelope reconstruction mechanisms.

Keywords: speech, hearing loss, envelope reconstruction, amplitude modulation, frequency modulation

Correspondence to: Christian Lorenzi · Equipe Audition (CNRS, Universite Paris Descartes, Ecole normale superieure), Institut d'Etude de la Cognition · Ecole normale superieure, Paris Sciences et Lettres · 29 rue d'Ulm, 75005 Paris, France. Telephone: +33 144322674; fax: +33 144322680; email: lorenzi@ens.fr

INTRODUCTION

A wide range of studies suggest that amplitude-modulation (AM) cues play an important role in speech intelligibility (e.g., Drullman 1995; Shannon et al. 1995). However, additional studies conducted with normal-hearing (NH) or hearing-impaired (HI) listeners have suggested that frequency-modulation (FM) cues may also be important for speech intelligibility when speech cues are severely degraded by background noise, interfering voices, filtering, infinite compression, peak clipping, or gating (e.g., Zeng et al. 2005; Gilbert and Lorenzi 2006, 2010; Lorenzi et al. 2006, 2009; Drennan et al. 2007; Ardoint and Lorenzi 2010; Gnansia et al. 2009, 2010; Hopkins et al. 2008, 2010; Ardoint et al. 2011; Eaves et al. 2011).

Psychoacoustical investigations (e.g., Saberi and Hafter 1995; Moore and Sek 1996) proposed that the perception of FM cues is mediated by two early auditory mechanisms: (1) a purely temporal mechanism based on phase-locking in auditory-nerve fibers, extracting so-called temporal fine-structure cues from FM, and (2) a second mechanism based on an AM (or so-called temporal envelope) reconstruction process. In this case, the differential attenuation of cochlear filtering converts the frequency excursions of FM into AM fluctuations at the output of cochlear filters.

In agreement with the second mechanism, psychoacoustical, electrophysiological, and modeling studies showed that the narrow-band speech AM patterns can be reconstructed from the broadband FM components as a result of cochlear filtering (Ghitza 2001; Zeng et al. 2004; Gilbert and Lorenzi 2006; Sheft et al. 2008; Heinz and Swaminathan 2009; Ibrahim and Bruce 2010; Swaminathan and Heinz 2012; see also Apoux et al. 2011). Gilbert and Lorenzi (2006) demonstrated that recovered AM cues contribute substantially to the identification of vocoded syllables retaining FM cues only when a wide analysis filter bandwidth is used to generate the vocoded signals. More precisely, their results showed that recovered AM cues played a major role in the identification of FM-vocoded syllables when the bandwidth of analysis filters used to vocode speech was greater than about four times the bandwidth of a normal cochlear filter (i.e., >4 ERB_N, where ERB_N stands for the average equivalent rectangular bandwidth of the cochlear filter as determined using young normally hearing listeners tested at moderate sound levels; cf. Glasberg and Moore 1990; Moore 2007). Consistent with these behavioral findings, the correlation between the original and the recovered AM components computed at the output of a simulated cochlear filterbank was also found to increase with the bandwidth of analysis bands (Gilbert and Lorenzi 2006; Sheft et al. 2008).

Thus, when only FM speech cues are presented to NH listeners within broad frequency bands, the recovered temporal-envelope information may serve as a strong cue for speech identification and help maintain robust speech intelligibility. HI listeners typically show normal or near-normal ability to use AM speech cues (e.g., Turner et al. 1995; Baskent 2006; Lorenzi et al. 2006, 2009). However, they may not be able to recover temporal-envelope information from FM as well as NH listeners do because cochlear filters tend to broaden with the amount of hearing loss (see Moore 2007, for a review). This prediction is supported qualitatively by the results of a modeling study reporting simulations of a computational model of auditory-nerve fibers' responses to a vocoded speech signal retaining broadband FM cues only (Heinz and Swaminathan 2009). Simulations obtained with an auditory-nerve fiber model tuned to a low center frequency (550 Hz) showed that cochlear envelopes recovered from FM cues extracted from one, two, or four broad frequency bands spanning the range 80–8820 Hz may be reduced up to about 40 % following a 30-dB hearing loss due to selective outer-hair-cell damage.[1] The model tuning curves were based on those obtained for NH cats. Modeling work performed by Ibrahim and Bruce (2010) indicated that the model of cat AM tuning underestimates envelope recovery in humans, for which frequency tuning was estimated to be two to three times sharper than cats (Recio et al. 2002; Shera et al. 2002). For this reason, the simulation data of Heinz and Swaminathan (2009) cannot be taken as precise predictions of the effect of cochlear damage on envelope recovery in humans. Nevertheless, they suggest that human subjects with cochlear hearing loss and the associated reduction in frequency selectivity should show poorer-than-normal vocoded-speech identification performance if a wide analysis filter bandwidth is used to generate the vocoded speech signals retaining FM cues only. The present study aimed to test this hypothesis and quantify precisely the potential deficits in speech-envelope reconstruction by measuring the ability of human listeners with mild to moderate hearing loss to identify vocoded speech signals retaining broadband FM cues only.

METHODS

The first experiment assessed the capacity of NH and HI listeners to identify nonsense syllables processed to retain FM cues only within three, ~7–8-ERB_N-wide analysis bands spanning the range 65–3,645 Hz.

A second experiment was conducted on NH listeners to assess the capacity to identify the FM-vocoded syllables on the sole basis of the putative envelope cues recovered at the output of auditory filters with normal (1-ERB_N wide) or broader-than-normal (2- or 4-ERB_N wide) bandwidths.

---

[1] A neural cross-correlation coefficient (ρenv) was used to predict the similarity of temporal-envelope cues between single auditory-nerve (AN) fiber spike-train responses to an intact speech token and corresponding FM-vocoded version. Recovery of speech envelope cues was demonstrated by a large envelope correlation ρ_{env} between the neural responses to the original speech token and to the corresponding FM-vocoded version. Simulation of a 30-dB hearing loss due to selective outer-hair-cell damage in the cochlear model (which corresponds to a broadening in cochlear tuning by a factor 1.4–2.3) resulted in a 40 % reduction of the ρ_{env} metric quantifying the amount of envelope reconstruction (see Heinz and Swaminathan 2009 for more details).

Participants

Forty NH and 41 HI listeners participated. The study was approved by CPP Ile de France (07018—ID RCB, 2007-A00343-50). Audiograms were measured for each listener using a Madsen Aurical +3.08 audiometer and TDH39 headphones in a single-wall sound-attenuating booth for the tested ear (right ear for all listeners). Pure-tone average (PTA) estimates were calculated for audiometric frequencies between 0.125 and 4 kHz for each individual NH and HI listener.

The listeners with normal hearing showed audiometric thresholds not exceeding 20 dB hearing level (HL) between 0.125 and 4 kHz. Their mean audiogram is shown in Figure 1 (dashed line). The listeners with normal hearing were considered as a single NH group according to the current experimental design. However, they were divided into two groups—young [n=16; mean age=24 years; standard deviation (SD)= 3 years; range=19–29 years; mean PTA=4.4 dB HL] and older (n=24; mean age=46 years; SD=10 years; range=31–63 years; mean PTA=6.5 dB HL) listeners —for further statistical analysis because of differences in their audiometric configuration. Young NH listeners showed audiometric thresholds not exceeding 20 dB HL up to (and including) 8 kHz. Fifteen among 24 older NH listeners showed audiometric thresholds not exceeding 20 dB HL up to (and including) 8 kHz; the remaining nine older NH listeners showed higher audiometric thresholds at 6 and 8 kHz (range=−5 to 50 dB HL).
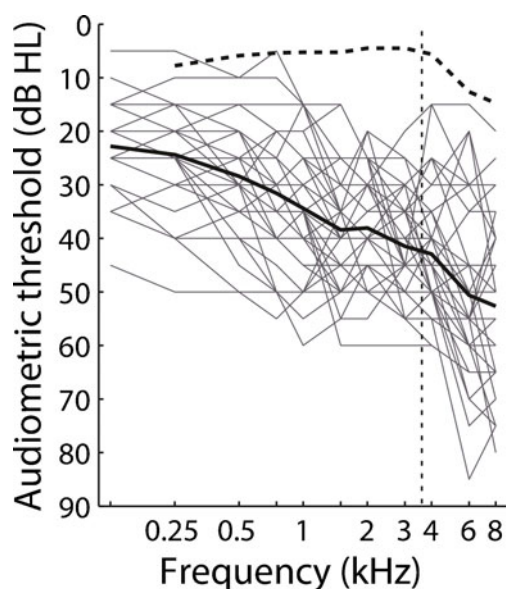
The listeners with hearing loss were aged between 28 and 70 years (n=41; mean age=58 years). Their audiograms [individual (gray lines) and mean data (bold line)] are shown in Figure 1. Air-conduction, bone-conduction, and impedance audiometry for the HI listeners were consistent with sensorineural impairment. HI listeners showed audiometric thresholds not exceeding 60 dB HL between 0.125 and 4 kHz. Their PTA calculated for audiometric frequencies between 0.125 and 4 kHz ranged between 20 and 48 dB HL (mean PTA=35 dB HL; SD=6 dB HL), indicating that the hearing loss was mild to moderate (Goodman 1965).

All speech stimuli were delivered monaurally to their right ear via Sennheiser HD280 headphones. For NH listeners, the presentation level of the stimuli was 65 dB SPL. Overall levels were calibrated using a Bruel & Kjaer (2250) sound level meter and ear simulator (B&K 4153) complying with IEC 60318-1. For each HI listener, presentation level was adjusted individually according to the "Cambridge formula" (Moore and Glasberg 1998), which corresponds to a modified half-gain linear amplification rule aiming to limit the upward spread of masking produced by amplification. Maximum presentation level never exceeded 90 dB SPL.

Speech identification tasks

Speech material. Forty-eight nonsense vowel-consonant-vowel-consonant-vowel (VCVCV) stimuli were spoken in quiet by a French female talker (mean VCVCV duration= 1,272 ms; standard deviation=113 ms), recorded digitally via a 16-bit A/D converter at a 44.1-kHz sampling frequency and equalized in root-mean-square (rms) power. These VCVCV stimuli consisted of three recordings of 16 /aCaCa/ utterances (C=/p, t, k, b, d, g, f, s, ʃ, v, z, j, m, n, r, l/). The fundamental frequency of the female voice was estimated to be 219 Hz using the algorithm of de Cheveigne and Kawahara (2002).

Speech-processing conditions. The original speech signals were submitted to three different processing schemes. Stimuli processed using the first scheme (referred to as "3-band AM+FM") contained AM and FM information from the speech waveform. Stimuli processed using the second scheme (referred to as "3-band FM") contained speech information in their FM only, but temporal-envelope cues were potentially recoverable at the output of auditory filters. Stimuli were finally generated using a third scheme (referred to as "Recov") so as to force listeners to identify consonants primarily on the basis of the envelope cues recovered at the output of simulated auditory filters from the 3-band FM stimuli. Three-band AM +FM and 3-band FM stimuli were presented for identification to NH and HI listeners. Recov stimuli were presented for identification to NH listeners only.

FIG. 1. Pure-tone audiograms for the tested (i.e., right) ear of the 41 HI listeners. The *bold continuous line* shows the mean audiogram of the 41 HI listeners. The *bold dashed line* shows the mean audiogram of the NH listeners. The *vertical thin dashed line* shows the high cutoff frequency (3.6 kHz) of the speech stimuli used in the present experiment.

Each /aCaCa/ signal was initially bandpass filtered using fourth-order gammatone filters (Patterson 1987) to generate the 3-band AM+FM and 3-band FM stimuli. For both processing conditions, the filterbank consisted of three, ~7–8-ERB_N-wide contiguous analysis bands spanning the range of 65 to 3,645 Hz.[2] The three analysis bands were centered on 263, 940, and 2,535 Hz, respectively. Their bandwidths were 395, 936, and 2,220 Hz, respectively. The three analysis bands encompassed the energy of the fundamental frequency (band 1), the first and second formant (band 2), and the third formant (band 3) of the French vowel /a/ used in this study, respectively. A speech-shaped noise (SSN) was generated whose spectrum matched the average spectrum of the whole set of /aCaCa/ stimuli. This SSN was passed through a highpass Butterworth filter (cutoff frequency=3,645 Hz, slope=108 dB/octave) and added to the processed speech stimuli (see below) to prevent the use of information from the transition band (Warren et al. 2004). The overall level of the highpass filtered noise was 12 dB below the overall level of the speech stimuli. The filtered noise was gated on and off with the target speech item, and shaped using a raised-cosine function with 10-ms rise/fall times. A different sample of the filtered noise was used for each presentation.

Three-band AM+FM condition. The Hilbert transform was applied to the bandpass-filtered speech signal in order to decompose the latter into its AM (modulus of the Hilbert analytic signal) and FM (cosine of the argument of the Hilbert analytic signal). For each band, the AM and FM components were multiplied. Impulse responses were peak-aligned for the AM (using a group delay of 16 ms) and the FM components across frequency channels (Hohmann 2002). The modulated signals were finally weighted and summed over the three frequency bands. The weighting compensated for imperfect superposition of the bands' impulse responses at the desired group delay. The weights were optimized numerically to achieve a flat frequency response (Hohmann 2002). The stimuli were equated in root mean square (rms) power.

Three-band FM condition. The Hilbert transform was applied to the bandpass-filtered speech signal in

order to decompose the latter into its AM and FM. The AM component was modified by replacing all amplitudes above a level of 20 dB below the band rms level by +1 and all amplitudes below that level by 0. This was achieved to avoid the amplification of low-level recording noise to the same level as the target speech (a signal-processing artifact that might have reduced speech intelligibility in previous studies conducted with HI listeners; see Hopkins et al. 2010). The modified AM was then lowpass-filtered (cutoff frequency=128 Hz; 56 dB/octave rolloff) using a zero-phase Butterworth filter to smooth fast transitions in the AM, and multiplied with the FM component. As for the AM+FM condition, impulse responses were peak-aligned for the AM and the FM components across frequency channels. The modulated signals were finally weighted and summed over the three frequency bands. The resulting stimuli were equated in rms power.

The effects of speech processing and degraded auditory frequency selectivity are illustrated in Figures 2 and 3 for a single speech utterance (/ababa/). The highpass-filtered SSN was not added to the processed speech signals for clarity. The left and right panels of Figure 2 show the waveform of the /ababa/ utterance in the 3-band AM+FM and 3-band FM conditions, respectively. Figure 2 shows that speech processing for the 3-band FM condition behaves like infinite peak clipping or like a very fast, multi-channel compressor with an infinite compression ratio (cf. Licklider and Pollack 1948; Gilbert and Lorenzi 2006). Figure 3 displays the temporal envelopes computed by a model of early auditory processing in response to the 3-band AM+FM (top panel) and 3-band FM (bottom panels) /ababa/ stimuli. Temporal envelopes were computed at the output of 80 gammatone filters tuned between 80 and 3758 Hz, via half-wave rectification and lowpass filtering at 20 Hz (Butterworth filter, rolloff=12 dB/octave).
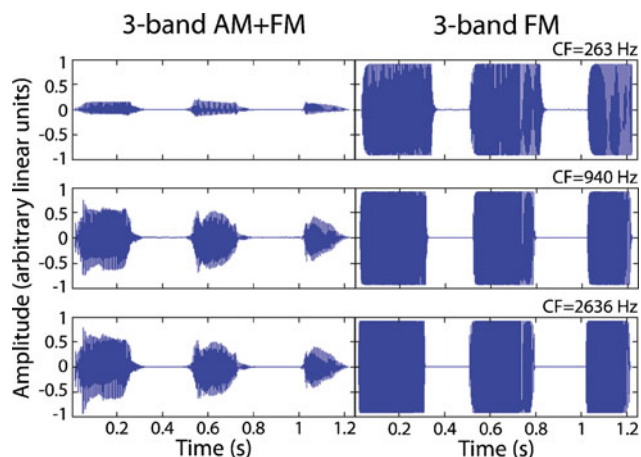

FIG. 2. Waveforms of a single /ababa/ utterance in the 3-band AM+FM (*left panel*) and 3-band FM (*right panel*) conditions. Waveforms are shown for each analysis band of center frequency CF.

---

[2] More precisely, speech sounds were initially split into four frequency bands spanning the range 65–8,958 Hz, and the fourth band centered on 6,325 Hz was omitted. Gilbert and Lorenzi (2006) and Heinz and Swaminathan (2009) used a comparable frequency range to generate four-band vocoded stimuli: 80–8,020 Hz in Gilbert and Lorenzi (2006) and 80–8,820 Hz in Heinz and Swaminathan (2009). As a consequence, the "4-band vocoder" conditions were somewhat comparable in Gilbert and Lorenzi (2006), Heinz and Swaminathan (2009), and in the present study.
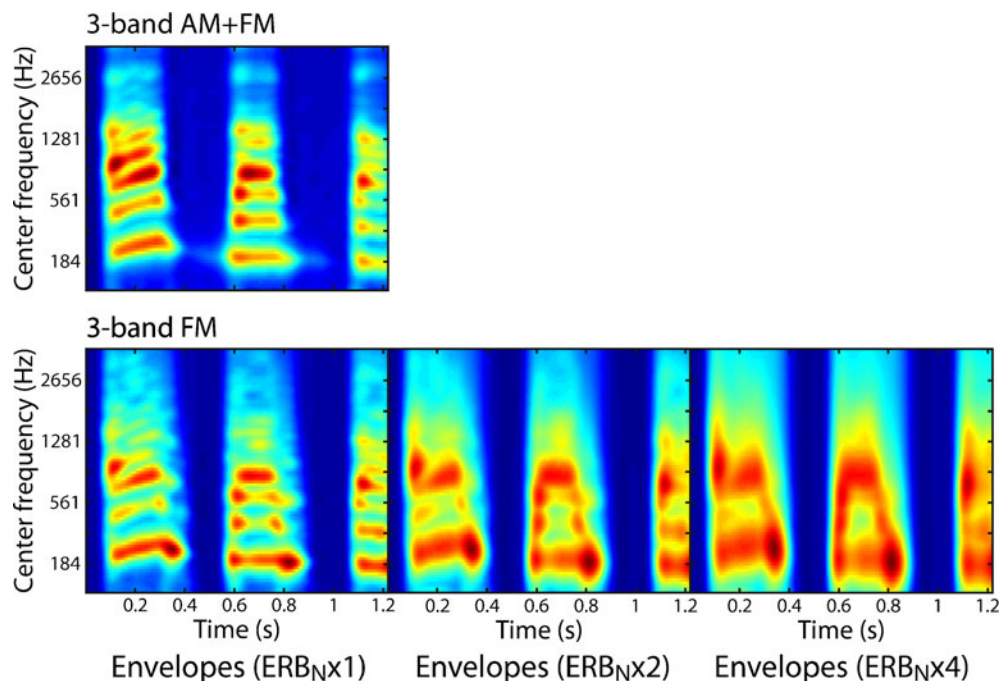
FIG. 3. Temporal envelopes of a single /ababa/ utterance in the 3-band AM+FM (*top panel*) and 3-band FM (*bottom panels*) conditions. Temporal envelopes are computed at the output of 80 1-ERB_N-wide (*top panel*, and *left panel*, *bottom row*), 2-ERB_N-wide (*middle panel*, *bottom row*), and 4-ERB_N-wide (*right panel*, *bottom row*) gammatone filters. In each panel, the center frequency (CF) of the gammatone filter is plotted on the ordinate, and time is plotted on the abscissa. A color scale ranging from blue to red is used, with a higher value for the red component.

Temporal envelopes were computed using 1-ERB_N-wide (top panel and left panel, bottom row) gammatone filters to simulate normal auditory filters. The model also simulated the effects of broadening auditory filters by factors of 2 (middle panel, bottom row) and 4 (right panel, bottom row), which is representative of mild/moderate and severe hearing loss, respectively (e.g., Tyler et al. 1984; Glasberg and Moore 1986; Dubno and Dirks 1989; Peters and Moore 1992; Stone et al. 1992; see also Moore 2007 for a review). Figure 3 shows that the original temporal-envelope speech cues (top panel) are recovered from FM cues at the output of simulated auditory filters when the latter are 1-ERB_N wide (left panel, bottom row). The recovered temporal-envelope pattern shown in the left panel of the bottom row displays the harmonic structure, the formants and their transitions, and the phonetic/syllabic segments of the /ababa/ signal. As expected, the middle and right panels of the bottom row show that the temporal-envelope cues recovered at the output of simulated auditory filters are substantially smeared when the bandwidth of auditory filters is broadened by a factor ranging between 2 and 4. In this case, the formants and the syllabic segments of the /ababa/ signal are only signaled by gross (recovered) envelope features encompassing broad frequency regions.

Recov condition. The 3-band FM signals were passed through a bank of 32 gammatone auditory filters, each 1-, 2-, or 4-ERB_N wide with center frequencies ranging from 80 to 8,583 Hz, and spaced along an ERB scale. In each band, the temporal envelopes were extracted using the Hilbert transform and lowpass-filtered using a zero-phase Butterworth filter. The envelopes were lowpass-filtered at $ERB_N/2$ for analysis filters whose bandwidths were greater than 128 Hz. Otherwise, the envelopes were lowpass-filtered at 64 Hz. These envelopes were then used to amplitude-modulate sine waves having the same frequencies as the original center frequencies of the auditory filters, but with random starting phase. In order to avoid delays in envelopes induced by filtering, impulse responses were peak-aligned for the envelope (using a group delay of 16 ms) and the carrier across frequency channels (Hohmann 2002). This envelope alignment was applied to the different stimuli. The modulated signals were finally weighted and summed over the 32 frequency bands. The weighting compensated for imperfect superposition of the bands' impulse responses at the desired group delay. The weights were optimized numerically to achieve a flat frequency response (Hohmann 2002; see also Gnansia et al. 2008, 2009, 2010 for further details).

This experimental condition was included to assess the capacity of NH listeners to identify the FM-vocoded speech sounds on the basis of the envelope

cues recovered at the output of auditory filters with normal (1-ERB$_N$ wide) or broader-than-normal (2- or 4-ERB$_N$ wide) bandwidths. As indicated above, the PTA calculated for audiometric frequencies between 0.125 and 4 kHz ranged between 20 and 48 dB HL, indicating that the hearing loss of the present HI listeners was globally mild to moderate (Goodman 1965). However, at each audiometric frequency, pure-tone sensitivity varied widely across individuals. At low frequencies (≤0.5 kHz) where envelope reconstruction was shown to be most prominent in previous modeling studies (e.g., Sheft et al. 2008; Heinz and Swaminathan 2009), audiometric thresholds ranged between 5 and 50 dB HL across HI listeners. At mid-frequencies (0.75–3 kHz), audiometric thresholds ranged between 5 and 60 dB HL across HI listeners. Broadening factors by 2 and 4 (i.e., 2- or 4-ERB$_N$-wide gammatone filters) were thus chosen to match the average auditory-filter broadening reported in the literature (e.g., Moore 2007) for HI listeners with the same range of audiometric thresholds as in here (5–60 dB HL). Note that a broadening of auditory filters by a factor of 2 should be considered even for mild levels of hearing loss. For the low-frequency region and for hearing losses between 5 and 30 dB HL, Moore (2007) reported a maximum broadening by a factor of 1.7 at 0.4 and 0.8 kHz; Santurette and Dau (2007) reported a maximum broadening by a factor of 1.4 at 0.5 kHz, and Strelcyk and Dau (2009) reported a maximum broadening by a factor of 1.4 at 0.75 kHz. For the mid-frequency region, Badri et al. (2011) reported a maximum broadening by a factor of 1.7 at 2 kHz.

Testing procedure

Speech identification was conducted in a sound-attenuating booth. All listeners were tested in a single session lasting between 1 and 2 h. This session was divided into several blocks, each corresponding to a given experimental condition. Before the actual testing session, the listeners were familiarized with speech material by listening to each original (i.e., unprocessed) logatome three times. During each block, the 48 processed /aCaCa/stimuli were presented in random order. On each trial, a single /aCaCa/ stimulus was presented to the listener with the task to identify the /aCaCa/ stimulus among 16 alternatives displayed on a computer screen. The alternatives were letter strings corresponding to the nonsense syllables (e.g., "ababa", "adada", etc.). Listeners responded by selecting a given letter string on the screen with a computer mouse. Response time was not limited, and no feedback was provided to listeners. Percent correct scores were corrected for chance level (6.25 %). The correction used was: PC corrected for chance $= [(PC$ raw $-$ Chance level$)/(100-$

Chance level$)] \times 100$, where PC stands for percent correct scores (Baskent 2006). In the corrected scores, a performance at 0 % represents a performance at chance level. The corrected scores were then transformed to rationalized arcsine units (RAU; Studebaker 1985). RAU scores are similar to percent correct scores for values of the latter from 15 % to 85 %, but RAU scores are "stretched" outside that region (reducing the effects of the bounded percent-correct scale). Speech identification was always assessed first for 3-band AM+FM speech stimuli in order to familiarize all listeners with the experimental procedure in the less distorted speech-processing condition. Then, speech identification was assessed for 3-band FM speech stimuli. Finally, speech identification was assessed for Recov speech stimuli. However, for each speech-processing condition, HI listeners were tested in random order for the two levels of stimulus presentation. In the Recov condition, NH listeners were tested in random order for the three bandwidths of analysis filters (1-, 2-, and 4-ERB$_N$-wide filters). For each HI listener, testing lasted about 15 min per experimental condition.

RESULTS

Experiment 1

The left and middle panels of Figure 4 show individual (filled symbols) and mean (horizontal bar in each box) identification scores expressed in RAU for the NH (circles) and HI (squares) listeners. The identification data are shown for the 3-band AM+FM (left panel) and 3-band FM (middle panel) conditions. NH data are shown for stimuli presented at a level of 65 dB SPL. HI data are shown for stimuli presented at a level of 65 dB SPL (condition "HI") or for stimuli amplified linearly using the Cambridge formula (condition "HI+").

Figure 4 shows that for most experimental conditions, HI data were more scattered than NH data. When stimuli were presented at 65 dB SPL, the identification scores of two HI listeners were extremely poor and slightly greater than chance level (−16 RAU) for both speech-processing conditions. At both presentation levels, the identification scores of 15 NH and five HI listeners were at ceiling level (116 RAU) in the 3-band AM+FM speech-processing condition. Only the 3-band FM condition appeared totally immune to ceiling *and* floor effects for both NH and HI listeners when stimulus audibility was controlled via linear amplification.

For each group of listeners and presentation level, average identification scores decreased by 9 to 30 RAU when the AM component was removed within each analysis band. Still, average identification scores remained above 50 RAU (and thus, well above chance level; here −16 RAU) across groups and presentation
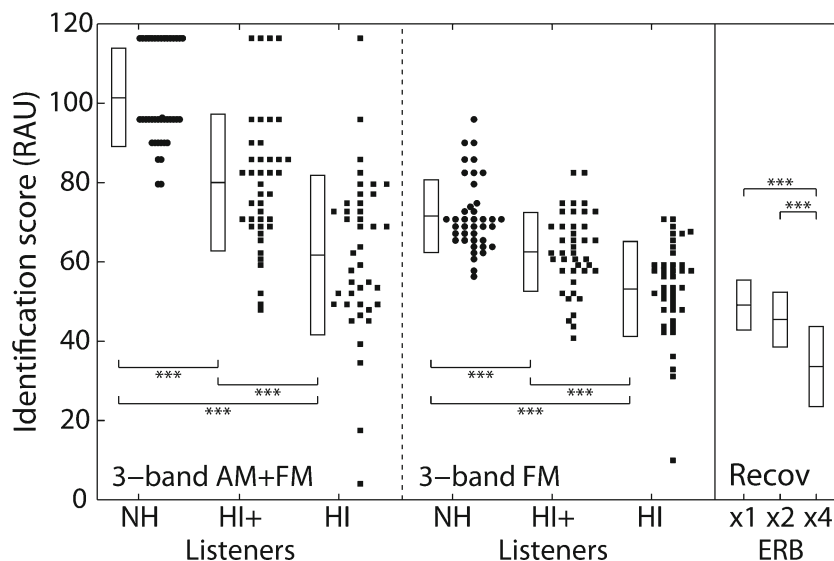
FIG. 4. Individual (*filled symbols*) and mean (*horizontal bar* in each box) identification scores (in RAU). *Open boxes* represent ±1 standard deviation about the mean for each condition. Percent-correct identification scores were corrected for chance level (6.25 % correct) and transformed to rationalized arcsine units (RAU). Floor and ceiling values correspond to −16 and 116 RAU, respectively. *Left* and *middle* panels: data are shown for the NH (*circles*) and HI (*squares*) listeners in the 3-band AM+FM (*left panel*) and 3-band FM (*middle panel*) conditions. NH data are shown for stimuli presented at a level of 65 dB SPL. HI data are shown for stimuli presented at a level of 65 dB SPL (*HI*) or amplified linearly (*HI+*). Asterisks (***) indicate that all comparisons of speech scores (NH/HI, NH/HI+, and HI/HI+) were significant at the 0.001 level. *Right panel*: mean identification scores obtained by NH listeners in the Recov condition, that is for 3-band FM speech passed through a tone-excited vocoder using 1-ERB_N-wide analysis filters ($ERB_N \times 1$), 2-ERB_N-wide analysis filters ($ERB_N \times 2$), or 4-ERB_N-wide analysis filters ($ERB_N \times 4$). Again, asterisks (***) indicate that comparisons of speech scores ($ERB_N \times 1/ERB_N \times 4$ and $ERB_N \times 2/ERB_N \times 4$) were significant at the 0.001 level.

levels after AM removal. For NH listeners, the removal of AM reduced average identification scores by 30 RAU. For HI listeners, the removal of AM had a greater deleterious effect when stimuli were amplified (17 RAU) than when they were presented at 65 dB SPL (9 RAU). The effect of removing AM was larger for NH listeners because of the stretching effect produced by the RAU transform on percent-correct identification scores (15 NH listeners being at ceiling performance for the 3-band AM+FM condition). For each speech-processing scheme, average scores measured for HI listeners increased when stimuli were submitted to linear amplification. This increase was two times greater for 3-band AM+FM stimuli (18 RAU) than for 3-band FM stimuli (9 RAU). However, for each presentation level, average scores of HI listeners remained poorer than normal (by 9 to 40 RAU) in each speech-processing condition. In the 3-band FM condition, the difference between NH and HI listeners' mean identification scores was two times smaller when stimuli were amplified (9 RAU) than when they were presented at 65 dB SPL (18 RAU).

Student *t* tests[3] (with Bonferroni correction) were conducted on the RAU scores obtained by NH and HI

listeners using 3-band AM+FM stimuli. These analyses showed significant differences between RAU scores when comparing (1) NH and HI listeners' data at 65 dB SPL ($p<0.001$), (2) NH listeners' data measured at 65 dB SPL and HI listeners' data measured with amplified stimuli ($p<0.001$), and (iii) HI listeners' data measured with and without amplification ($p<0.001$).

Similar Student *t* tests (with Bonferroni correction) were then conducted on the RAU scores obtained by NH and HI listeners using 3-band FM stimuli (these two statistical analyses were conducted separately because listeners were always tested first with 3-band AM+FM stimuli). Again, statistical analyses showed significant differences between RAU scores when comparing (1) NH and HI listeners' data at 65 dB SPL ($p<0.001$), (2) NH listeners' data measured at 65 dB SPL and HI listeners' data measured with amplified stimuli ($p<0.001$), and (3) HI listeners' data measured with and without amplification ($p<0.001$).

In summary, both NH and HI listeners demonstrated on average a clear ability to identify consonants on the sole basis of the broadband FM speech cues. The ability of HI listeners to identify 3-band FM speech stimuli was significantly poorer than normal with or without linear amplification of speech stimuli. However, the overall deficit shown by HI listeners with amplified 3-band FM speech stimuli was modest,

---

[3] All statistical tests were performed using the R software (R Project).

amounting to 9 RAU. Interestingly, the effect of amplification was greater for 3-band AM+FM than for 3-band FM stimuli, suggesting that the ability to identify consonants processed to retain FM cues only was less affected by a change in stimulus audibility than the ability to identify consonants with intact modulation components. However, this result should be treated with caution because all listeners were tested first with 3-band AM+FM stimuli.

Experiment 2

The right panel of Figure 4 shows the mean scores (horizontal bar in each box) obtained by NH listeners in the Recov condition. For each Recov condition, the area around the mean identification score indicates ±1 standard deviation about the mean.

Consistent with previous work (e.g., Gilbert and Lorenzi 2006), NH listeners could identify 3-band FM speech stimuli on the basis of recovered envelope cues well above chance level. Mean identification scores were 49, 45, and 34 RAU for 1-, 2,- or 4-ERB_N-wide analysis filters, respectively. The mean identification score obtained with 1-ERB_N-wide analysis filters (49 RAU) was lower than the mean score obtained by NH listeners with 3-band FM stimuli (72 RAU), suggesting that factors other than peripheral envelope reconstruction may play a role in the identification of 3-band FM speech. Previous work suggests that the ability to use temporal fine-structure cues conveyed by patterns of phase locking in auditory-nerve fibers may have contributed to the identification of vocoded consonants (e.g., Lorenzi et al. 2006, 2009; Sheft et al. 2008; Heinz and Swaminathan 2009; Ardoint et al. 2010; however, see Swaminathan and Heinz 2012). On the other hand, it is likely that speech identification based on recovered envelope cues was underestimated with the present simulation approach because (1) human peripheral tuning may be two or more times sharper than previously estimated (Shera et al. 2002), such that envelope reconstruction may be stronger than originally assumed (Ibrahim and Bruce 2010), and (3) tone-vocoders degrade somewhat the transmission of envelope cues (Kates 2011). Thus, the *relative* pattern of performance obtained in the Recov condition (i.e., the effect of broadening analysis filters on identification scores) may be more relevant to the present investigation than absolute scores. Figure 4 shows that broadening analysis filters by a factor of 2 reduced identification scores by 4 RAU only, whereas broadening analysis filters by a factor of 4 reduced identification scores by 15 RAU. These data are in line with the modest deficits (9 RAU) observed for HI listeners with 3-band FM speech in the amplified condition (HI+) (as shown in the middle panel of Fig. 4).

A repeated measures analysis of variance (ANOVA) conducted on the RAU scores of the NH listeners in the three Recov conditions showed a significant effect of the broadening of the analysis filters [$F(2,78)=52$, $p<0.001$]. A post hoc analysis [Tukey HSD (honestly significant difference) test] showed that broadening analysis filters by a factor of 2 did not degrade significantly identification performance for NH listeners compared to the case where analysis filters were 1-ERB_N wide ($p=0.1$). However, broadening analysis filters by a factor of 4 reduced performance significantly (all $p<0.001$).

DISCUSSION

The ability to use recovered envelope cues

Mean identification scores measured for both NH and HI listeners using syllables processed to retain FM cues only were above 60 RAU—and thus, well above chance level (−16 RAU)—when stimulus audibility was controlled via linear amplification. This finding is consistent with the results of experiment 2 where mean identification scores based on the recovered envelope cues only were shown to range between about 35 and 50 RAU for NH listeners. These results suggest that, despite potentially important variations in auditory frequency selectivity (see Moore 2007), both NH listeners and listeners with mild-to-moderate hearing loss are able to maintain robust speech intelligibility in quiet by using temporal-envelope information recovered from the *broadband* FM speech cues. Note, however, that additional speech cues (such as temporal fine-structure cues) may have contributed to the identification of FM-vocoded speech identification for both NH and HI listeners.

The current study also showed that our listeners with mild-to-moderate cochlear hearing loss (mean PTA= 35 dB HL) were significantly worse than NH listeners in the 3-band FM speech condition. This suggested that their ability to reconstruct temporal-envelope cues was degraded by cochlear damage, consistent with the modeling work conducted by Heinz and Swaminathan (2009) with comparable vocoded-speech stimuli. The deficit shown by HI listeners was relatively small, amounting to 9 RAU for 3-band FM speech after stimulus amplification. Moreover, the effect of broadening analysis filters by a factor of 2 did not yield a substantial drop in 3-band FM speech identification performance for NH listeners (4 RAU). Altogether, these results indicate that HI listeners with mild-to-moderate hearing loss remain able to use temporal-envelope information recovered from broadband FM speech cues, although this capacity weakens because of degraded frequency selectivity.

A simulation study based on the correlational approach developed by Zeng et al. (2004), Gilbert and Lorenzi (2006), and Sheft et al. (2008) was conducted to better understand the limited effect of cochlear damage on the identification of 3-band FM speech stimuli. Quantitative estimates of the fidelity of temporal-envelope reconstruction were obtained by calculating the correlation between the temporal envelopes of the original speech stimuli and the 3-band FM speech stimuli at the output of 30 gammatone filters with center frequencies uniformly spaced along an ERB scale ranging from 80 to 3,500 Hz. In each band, the temporal envelopes were extracted using the Hilbert transform and lowpass filtering at $ERB_N/2$ Hz (using a 180th order finite impulse response filter). The correlation estimates were calculated for 1-, 2-, and 4-ERB_N-wide gammatone filters. For each condition, the correlation estimates were finally averaged across the entire set of 48 /aCaCa/ speech stimuli. A Fischer z transform was used before averaging correlations. The correlation estimates are reported in Figure 5 with results obtained with 1-, 2-, and 4-ERB_N-wide gammatone filters shown by bold, dashed, and dotted lines, respectively. The numbers in parentheses are the mean correlation across gammatone filters for each bandwidth of gammatone filters. Figure 5 shows that the mean correlation is relatively high (>0.5), and decreases only slightly as a function of gammatone filters' bandwidth. More precisely, the mean correlation is not affected by a broadening of gammatone filters by a factor of 2 (however, note that correlation estimates decrease slightly in the 200–500-Hz range). The mean correlation decreases only from 0.6 to 0.5 when the gammatone filters are broadened by a factor of 4. Most of the reduction in envelope reconstruction occurs for filters tuned at or close to the fundamental frequency of the female speaker (219 Hz). Similar results (not shown
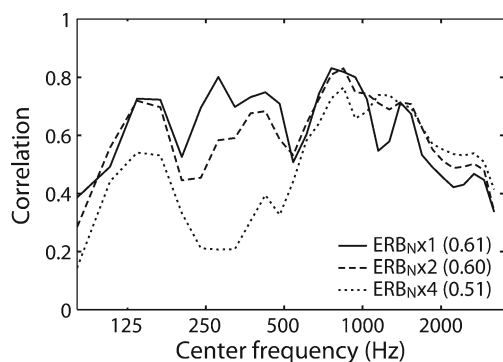
here) were obtained with other metrics assessing the fidelity of temporal-envelope transmission, such as the "depth-dependent" correlation index developed by Sheft et al. (2008) that takes into account changes in modulation depth at the output of gammatone filters (see Sheft et al. 2008 for further details).

These simulated data suggest that the loss of frequency selectivity typically associated with mild to moderate cochlear hearing loss should only have a modest deleterious effect on peripheral temporal-envelope reconstruction. This is in good agreement with the empirical data collected for most of our HI listeners. Thus, the differentiation process that converts FM excursions into temporal-envelope fluctuations at the output of cochlear filters appears hardly affected, if at all, by the reduction of the slope of auditory filters tuned to low/middle frequencies (when filters' bandwidth changes from 1 to 4 ERB_N, the low- and high-frequency slopes of gammatone filters decrease from about 80 to 10 dB/octave and from about 100 to 35 dB/octave, respectively). The global decrease in modulation depth of the reconstructed envelopes resulting from the broadening of auditory filters should have little impact on the reception of speech cues because auditory sensitivity to temporal-envelope cues is usually normal (or even better than normal) for HI listeners when audibility is carefully controlled for (Bacon and Gleitman 1992; Moore et al. 1992, 1996; Füllgrabe et al. 2003; Kale and Heinz 2010, 2012). The main deleterious effects of cochlear damage on the intelligibility of 3-band FM speech sounds would then correspond to (1) a reduction in the number of independent peripheral channels which transmit envelope information (see also Swaminathan and Heinz 2011) and (2) a reduction in the ability to discriminate and use envelope information for speech identification. Previous studies demonstrated that speech intelligibility in quiet is not affected by a reduction in frequency resolution by a factor of 2–4 (e.g., Baer and Moore 1993; Shannon et al. 1995; Baskent 2006). Other studies indicated that HI listeners may show poorer-than-normal AM discrimination abilities (Grant et al. 1998; Buss et al. 2004). This may partially explain why removing AM speech cues had a smaller detrimental effect on speech intelligibility for HI listeners compared to NH listeners in the present study. However, several studies showed that HI listeners are nearly as good as NH listeners at understanding AM-vocoded speech (e.g., Turner et al. 1995; Baskent 2006; Lorenzi et al. 2006). This suggests that HI listeners are as efficient as NH listeners at using complex temporal-envelope speech cues. Therefore, it is not surprising that the smearing of envelope cues illustrated in Figure 3 has only a small impact on consonant identification for the present HI listeners.

FIG. 5. Assessment of temporal-envelope reconstruction at the output of 1-ERB_N (*bold line*), 2-ERB_N (*dashed line*), and 4-ERB_N (*dotted line*)-wide gammatone filters. Mean correlation estimates between the original speech envelopes and the envelopes of the stimuli in the 3-band FM speech-processing conditions are shown as a function of gammatone filter center frequency. *Numbers* between *parentheses* correspond to the mean correlation estimate across channels in the respective gammatone-filter bandwidth condition.

The small deficit shown by HI listeners contrasts with the dramatic deficits reported by Lorenzi et al. (2006) and Ardoint et al. (2010) for HI listeners with mild-to-moderate cochlear hearing loss. The speech material and speech processing in these studies differ in three aspects that, together, could account for the difference in HI listeners' performance between previous and current investigations. First, VCV stimuli were used by Lorenzi et al. (2006) and Ardoint et al. (2010), whereas VCVCV stimuli were used in the current study. Thus, all listeners were provided with more redundant consonantal cues in the present study, and this may have limited the detrimental effects of speech processing (i.e., FM vocoding) on HI listeners' performance. Secondly, analysis bands with no speech information at a particular time were filled with distracting background sound in Lorenzi et al. (2006) and Ardoint et al. (2010). As a result, most FM speech sounded harsh and noisy. This may have posed a particular problem to the HI listeners who would have suffered more from masking between channels because of their broadened cochlear filters (Moore 2008). In the present study, such masking effects were strongly limited because FM information was extracted in each analysis band only when the band signal was above −20 re: RMS (that is, when the band signal was assumed to correspond to speech instead of low-level recording noise). Finally, speech sounds were split into 16 2-ERB$_N$-wide analysis filters in Lorenzi et al. (2006) and Ardoint et al. (2010), versus three 8-ERB$_N$-wide analysis filters in the present investigation. Gilbert and Lorenzi (2006) showed that for NH listeners, the amount of envelope reconstruction expected at the output of cochlear filters (and the ability to use these reconstructed envelope speech cues) increased strongly with the bandwidth of the FM-vocoder analysis bands (and thus, with the bandwidth of FM speech cues). It is reasonable to assume that the negative impact of reduced frequency selectivity caused by cochlear damage on the amount of envelope reconstruction was smaller when envelope cues were reconstructed from broadband FM cues (as in the case of a 3-band FM vocoder) rather than from narrowband FM cues (as in the case of a 16-band FM vocoder). However, it is difficult to reconcile this explanation with the outcome of the simulation studies conducted by Heinz and Swaminathan (2009) and Ardoint et al. (2010), showing very limited effects of reduced frequency selectivity on envelope reconstruction for 16-band FM speech stimuli. Two explanations may be proposed: (1) the contribution of (neural) temporal fine-structure cues to the identification of FM vocoded speech sounds may vary inversely with the availability of reconstructed envelope cues (in that case, a limited capacity to encode and/or use temporal fine-structure cues may explain

the dramatic deficits reported for HI listeners with 16-band FM-speech stimuli; Buss et al. 2004; Lorenzi et al. 2006; Moore 2008), or (2) the contribution of reconstructed envelope cues to FM-speech identification has been underestimated when FM cues are extracted from 16 narrow analysis bands (as suggested by Ibrahim and Bruce 2010).

Effects of age

The HI listeners who participated in the present study were globally older than the NH listeners: the age of NH listeners ranged between 19 and 63 years whereas the age of HI listeners ranged between 28 and 70 years. Correlation analyses performed between the individual RAU scores of the HI listeners in 3-band FM condition (in the amplified condition) and their age while factoring out the effect of the PTAs estimated below 4 kHz were not significant ($r=−0.02$, $p=0.92$), demonstrating that the identification deficits shown by HI listeners for 3-band FM stimuli could not be accounted for by aging per se. This result is consistent with the outcome of previous studies showing that neither auditory frequency selectivity (e.g., Peters and Moore 1992) nor the ability to discriminate complex temporal-envelope cues (e.g., Takahashi and Bacon 1992; Lorenzi et al. 2006) are strongly affected by aging.

Contribution of audibility

Correlation analyses performed between the individual RAU scores of the HI listeners in 3-band FM condition (in the amplified condition) and their PTAs estimated below 4 kHz while factoring out the effect of age were not significant ($r=−0.18$, $p=0.27$), demonstrating that the identification deficits shown by HI listeners for 3-band FM stimuli could not be accounted for by differences in PTA estimated below 4 kHz. Statistical analyses also showed that linear amplification had a much smaller effect on speech identification for 3-band FM stimuli than for 3-band AM+FM stimuli. Altogether, these results suggest that the reduced audibility shown by HI listeners was not responsible for their speech identification deficit in the amplified, 3-band FM condition.

Training effects

Previous work assessing the intelligibility of speech stimuli processed to retain only FM cues showed that the capacity to use the degraded speech cues was highly susceptible to training effects for NH listeners when a high frequency resolution (e.g., 16 or 32 bands) was used to generate vocoded stimuli (Lorenzi et al. 2006; Gilbert et al. 2007; Hopkins et al. 2010). In

the present study, all listeners were tested only once with a complete set of 3-band FM speech stimuli presented at a given sound level. Consequently, the capacity to use recovered envelope cues may have been underestimated in each group. An additional experiment was run on 10 among the 40 NH listeners to assess the extent to which the capacity to identify 3-band FM speech stimuli is affected by training. All NH listeners were tested again for ten successive blocks of 48, 3-band FM speech stimuli during a single session. An ANOVA conducted on the training data showed that consonant identification scores improved significantly by 9 RAU between the first and last training block [$F(10,90)=3$; $p<0.01$]. This small, although significant training effect suggests that the capacity of NH and HI listeners to use recovered envelope cues was not substantially underestimated when performing the current consonant identification task.

Clinical applications

The present speech identification test based on processed nonsense speech material (3-band FM syllables) may lead to interesting clinical applications. This test (1) is easy and relatively fast to administer, (2) does not require training, (3) yields high levels of performance in normal control listeners, (4) is relatively immune to floor and ceiling effects in normal control listeners and listeners with a mild-to-moderate hearing loss, and (5) is less influenced by reduced audibility than a comparable speech test using unprocessed speech material. The outcome of this study indicates that poorer-than-normal frequency selectivity impairs—although modestly—the capacity to recover important envelope information from broadband FM speech cues. This test may thus be of use for the fast diagnosis of supra-threshold auditory deficits responsible for impoverished speech perception in daily situations where AM cues are degraded by nonlinear distortions (e.g., clipping, compression; Houtgast and Steeneken 1985).

CONCLUSIONS

The present study compared the ability of a large cohort of normal-hearing and hearing-impaired listeners to identify nonsense syllables processed to retain FM cues only within three broad frequency bands.

1. Consistent with previous modeling work, hearing-impaired listeners showed significantly poorer identification scores than normal-hearing listeners irrespective of presentation level.
2. The detrimental effect of cochlear hearing loss was relatively modest, amounting to 9 RAU when

speech stimuli were amplified to compensate for reduced audibility.
3. Overall, the hearing-impaired data were consistent with a poorer-than-normal ability to reconstruct temporal-envelope speech cues resulting from a broadening of cochlear filters by a factor ranging from 2 to 4.

These results indicate that temporal-envelope reconstruction from broadband FM is an important, early auditory mechanism contributing to the robust perception of speech sounds in degraded listening conditions, as also suggested by recent work by Swaminathan and Heinz (2012). These results also suggest that most people suffering from mild to moderate cochlear hearing loss can make efficient use of reconstructed envelope cues despite degradations in frequency selectivity. Still, our results suggest that poorer-than-normal frequency selectivity impairs somewhat temporal-envelope reconstruction mechanisms. For hearing-impaired listeners showing the worst capacity to use recovered envelope cues, poor speech perception is anticipated in listening situations where AM speech cues are attenuated (for instance, by peak clipping or amplitude compression). The current study shows directions to develop a clinical test that may prove useful for the fast diagnosis of supra-threshold auditory deficits.

ACKNOWLEDGMENTS

REFERENCES

APOUX F, MILLMAN RE, VIEMEISTER NF, BROWN CA, BACON SP (2011) On the mechanism involved in the recovery of envelope information from temporal fine structure. J Acoust Soc Am 130:273–282

ARDOINT M, LORENZI C (2010) Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues. Hear Res 260:89–95

ARDOINT M, SHEFT S, FLEURIOT P, GARNIER S, LORENZI C (2010) Perception of temporal fine structure cues in speech with minimal envelope cues for listeners with mild-to-moderate hearing loss. Int J Audiol 49:823–831

ARDOINT M, AGUS T, SHEFT S, LORENZI C (2011) Importance of temporal-envelope speech cues in different spectral regions. J Acoust Soc Am Exp Lett 130:EL115–EL121

Bacon SP, Gleitman RM (1992) Modulation detection in subjects with relatively flat hearing losses. J Speech Hear Res 35:642–653

Badri R, Siegel JH, Wright BA (2011) Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms. J Acoust Soc Am 129:852–863

Baer T, Moore BCJ (1993) Effects of spectral smearing on the intelligibility of sentences in noise. J Acoust Soc Am 93:1229–1241

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 JW 3rd, Grose JH (2004) Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss. Ear Hear 25:242–250

de Cheveigne A, Kawahara H (2002) YIN, a fundamental frequency estimator for speech and music. J Acoust Soc Am 111:1917–1930

Drennan WR, Won JH, Dasika VK, Rubinstein JT (2007) Effects of temporal fine structure on the lateralization of speech and on speech understanding in noise. J Assoc Res Otol 8:373–383

Drullman R (1995) Temporal envelope and fine structure cues for speech intelligibility. J Acoust Soc Am 97:585–592

Dubno JR, Dirks DD (1989) Auditory filter characteristics and consonant recognition for hearing-impaired listeners. J Acoust Soc Am 85:1666–1675

Eaves JM, Kitterick PT, Summerfield AQ (2011) Benefit of temporal fine structure to speech perception in noise measured with controlled temporal envelopes. J Acoust Soc Am 130:501–507

Füllgrabe C, Meyer B, Lorenzi C (2003) Effect of cochlear damage on the detection of complex temporal envelopes. Hear Res 178:35–43

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, 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, Lorenzi C (2010) Role of spectral and temporal cues in restoring missing speech information. J Acoust Soc Am 128: EL294–EL299

Gilbert G, Bergeras I, Voillery D, 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

Glasberg BR, Moore BCJ (1986) Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. J Acoust Soc Am 79:1020–1033

Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data. Hear Res 47:103–138

Gnansia D, Jourdes V, Lorenzi C (2008) Effect of masker modulation depth on speech masking release. Hear Res 239:60–68

Gnansia D, Pean V, Meyer B, Lorenzi C (2009) Effects of spectral smearing and temporal fine structure degradation on speech masking release. J Acoust Soc Am 125:4023–4033

Gnansia D, Pressnitzer D, Pean V, Meyer B, Lorenzi C (2010) Intelligibility of interrupted and interleaved speech in normal-hearing listeners and cochlear implantees. Hear Res 14:46–53

Goodman A (1965) Reference zero levels for pure-tone audiometer. ASHA 7:262–263

Grant KW, Summers V, Leek MR (1998) Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners. J Acoust Soc Am 104:1051–1060

Heinz MG, Swaminathan J (2009) Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech. J Assoc Res Otolaryngol 10:407–423

Hohmann V (2002) Frequency analysis and synthesis using a Gammatone filterbank. Acust Acta Acust 88:433–442

Hopkins K, Moore BCJ, Stone MA (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

Hopkins K, Moore BCJ, Stone MA (2010) The effects of the addition of low-level, low-noise noise on the intelligibility of sentences processed to remove temporal envelope information. J Acoust Soc Am 128:2150–2161

Houtgast T, Steeneken HJM (1985) A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. J Acoust Soc Am 77:1069–1077

Ibrahim RA, Bruce IC (2010) Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues. In: Lopez-Poveda EA, Palmer AR, Meddis R (eds) The neurophysiological bases of auditory perception, chapter 40. Springer, New York, pp 429–438. doi:10.1007/978-1-4419-5686-6_40

Kale S, Heinz M (2010) Envelope coding in auditory nerve fibers following noise-induced hearing loss. J Assoc Res Otolaryngol 11:657–673

Kale S, Heinz M (2012) Temporal modulation transfer functions measured from auditory-nerve responses following sensorineural hearing loss. Hear Res 286:64–75

Kates JM (2011) Spectro-temporal envelope changes caused by temporal fine-structure modification. J Acoust Soc Am 129:3981–3990

Licklider JCR, Pollack I (1948) Effects of differentiation, integration and infinite peak clipping upon the intelligibility of speech. J Acoust Soc Am 20:42–51

Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BCJ (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, Debruille L, Garnier S, Fleuriot P, Moore BCJ (2009) Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. J Acoust Soc Am 125:27–30

Moore BCJ (2007) Cochlear hearing loss: physiological, psychological and technical issues. Wiley, Chichester

Moore BCJ (2008) 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 BCJ, Glasberg BR (1998) Use of a loudness model for hearing-aid fitting. I. Linear hearing aids. Br J Audiol 32:317–335

Moore BCJ, 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 BCJ, Shailer MJ, Schooneveldt GP (1992) Temporal modulation transfer functions for band-limited noise in subjects with cochlear hearing loss. Br J Audiol 26:229–237

Moore BCJ, Wotczak M, Vickers DA (1996) Effect of loudness recruitment on the perception of amplitude modulation. J Acoust Soc Am 100:481–489

Patterson RD (1987) A pulse ribbon model of monaural phase perception. J Acoust Soc Am 82:1560–1586

Peters RW, Moore BCJ (1992) Auditory filter shapes at low center frequencies in young and elderly hearing-impaired subjects. J Acoust Soc Am 91:256–266

Recio A, Rhode WS, Kiefte M, Kluender KR (2002) Responses to cochlear normalized speech stimuli in the auditory nerve of cat. J Acoust Soc Am 111:2213–2218

Saberi K, Hafter ER (1995) A common neural code for frequency- and amplitude-modulated sounds. Nature 374:537–539

Santurette S, Dau T (2007) Binaural pitch perception in normal-hearing and hearing-impaired listeners. Hear Res 223:29–47

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304

Sheft S, Ardoint M, Lorenzi C (2008) Speech identification based on temporal fine structure cues. J Acoust Soc Am 124:562–575

Shera CA, Guinan JJ Jr, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc Natl Acad Sci USA 99:3318–3323

Stone MA, Glasberg BR, Moore BCJ (1992) Simplified measurement of impaired auditory filter shapes using the notched-noise method. Br J Audiol 26:329–334

Strelcyk O, 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 GA (1985) A "rationalized" arcsine transform. J Speech Hear Res 28:455–462

Swaminathan J, Heinz MG (2011) Predicted effects of sensorineural hearing loss on across-fiber envelope coding in the auditory nerve. J Acoust Soc Am 129:4001–4013

Swaminathan J, Heinz MG (2012) Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. J Neurosci 32:1747–1756

Takahashi GA, Bacon SP (1992) Modulation detection, modulation masking, and speech understanding in noise in the elderly. J Speech Hear Res 35:1410–1421

Turner CW, Souza PE, Forget LN (1995) Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. J Acoust Soc Am 97:2568–2576

Tyler RS, Hall JW, Glasberg BR, Moore BCJ, Patterson RD (1984) Auditory filter asymmetry in the hearing impaired. J Acoust Soc Am 76:1363–1368.

Warren RM, Bashford JA Jr, Lenz PW (2004) Intelligibility of bandpass filtered speech: steepness of slopes required to eliminate transition band contributions. J Acoust Soc Am 115:1292–1295

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

Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargave A, Wei C, Cao K (2005) Speech recognition with amplitude and frequency modulations. Proc Natl Acad Sci USA 102:2293–2298