

Evaluating the role of spectral and envelope characteristics in the intelligibility advantage of clear speech

Jean C. Krause^{a)} and Louis D. Braid

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 4 March 2008; revised 10 February 2009; accepted 14 February 2009)

In adverse listening conditions, talkers can increase their intelligibility by speaking clearly [Picheny, M.A., *et al.* (1985). *J. Speech Hear. Res.* **28**, 96–103; Payton, K. L., *et al.* (1994). *J. Acoust. Soc. Am.* **95**, 1581–1592]. This modified speaking style, known as *clear speech*, is typically spoken more slowly than conversational speech [Picheny, M. A., *et al.* (1986). *J. Speech Hear. Res.* **29**, 434–446; Uchanski, R. M., *et al.* (1996). *J. Speech Hear. Res.* **39**, 494–509]. However, talkers can produce clear speech at normal rates (clear/normal speech) with training [Krause, J. C., and Braid, L. D. (2002). *J. Acoust. Soc. Am.* **112**, 2165–2172] suggesting that clear speech has some inherent acoustic properties, independent of rate, that contribute to its improved intelligibility. Identifying these acoustic properties could lead to improved signal processing schemes for hearing aids. Two global-level properties of clear/normal speech that appear likely to be associated with improved intelligibility are increased energy in the 1000–3000-Hz range of long-term spectra and increased modulation depth of low-frequency modulations of the intensity envelope [Krause, J. C., and Braid, L. D. (2004). *J. Acoust. Soc. Am.* **115**, 362–378]. In an attempt to isolate the contributions of these two properties to intelligibility, signal processing transformations were developed to manipulate each of these aspects of conversational speech independently. Results of intelligibility testing with hearing-impaired listeners and normal-hearing listeners in noise suggest that (1) increasing energy between 1000 and 3000 Hz does not fully account for the intelligibility benefit of clear/normal speech, and (2) simple filtering of the intensity envelope is generally detrimental to intelligibility. While other manipulations of the intensity envelope are required to determine conclusively the role of this factor in intelligibility, it is also likely that additional properties important for highly intelligible speech at normal rates remain to be identified.

© 2009 Acoustical Society of America. [DOI: 10.1121/1.3097491]

PACS number(s): 43.71.Ky, 43.66.Ts, 43.70.Gr, 43.70.Fq [MSS]

Pages: 3346–3357

I. INTRODUCTION

It has been well established that *clear speech*, a speaking style that many talkers adopt in difficult communication situations, is significantly more intelligible (roughly 17 percentage points) than conversational speech for both normal-hearing and hearing-impaired listeners in a variety of difficult listening situations (e.g., Picheny *et al.*, 1985). Specifically, the intelligibility benefit of clear speech (relative to conversational speech) has been demonstrated for backgrounds of wideband noise (e.g., Uchanski *et al.*, 1996; Krause and Braid, 2002) and multi-talker babble (e.g., Ferguson and Kewley-Port, 2002), after high-pass and low-pass filtering (Krause and Braid, 2003) and in reverberant environments (Payton *et al.*, 1994). Similar intelligibility benefits have also been shown for a number of other listener populations, including cochlear-implant users (Liu *et al.*, 2004), elderly adults (Helfer, 1998), children with and without learning disabilities (Bradlow *et al.*, 2003), and non-native listeners (Bradlow and Bent, 2002; Krause and Braid, 2003). These large and robust intelligibility differences are,

not surprisingly, associated with a number of acoustical differences between the two speaking styles. For example, clear speech is typically spoken more slowly than conversational speech (Picheny *et al.*, 1986; Bradlow *et al.*, 2003), with greater temporal envelope modulations (Payton *et al.*, 1994; Krause and Braid, 2004; Liu *et al.*, 2004) and with relatively more energy at higher frequencies (Picheny *et al.*, 1986; Krause and Braid, 2004). In addition, increased fundamental frequency variation (Krause and Braid, 2004; Bradlow *et al.*, 2003), increased vowel space (Picheny *et al.*, 1986; Ferguson and Kewley-Port, 2002), and other phonetic and phonological modifications (Krause and Braid, 2004; Bradlow *et al.*, 2003) are apparent in the clear speech of some talkers. Any of these differences between clear and conversational speech could potentially play a role in increasing intelligibility. However, the particular properties of clear speech that are responsible for its intelligibility advantage have not been isolated.

One approach that can be used to evaluate the relative contribution of various acoustic properties to the intelligibility advantage of clear speech is to develop signal processing schemes capable of manipulating each acoustic parameter independently. By comparing the intelligibility of speech before and after an individual acoustical parameter is altered, the contribution of that acoustic parameter to intelligibility

^{a)}Author to whom correspondence should be addressed. Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL. Electronic mail: jkrause@cas.usf.edu

can be quantified. Of course, it may be necessary to modify a combination of multiple parameters to reproduce the full benefit of clear speech; however, manipulation of single acoustic parameters is nonetheless a convenient tool for characterizing their relative importance (at least to the extent that intelligibility is a linear combination of these parameters). The effect of each parameter on intelligibility can be isolated, and its role in improving intelligibility can be evaluated systematically.

Several studies have used this systematic approach to analyze the role of speaking rate in the intelligibility of clear speech, each employing a different time-scaling procedure to alter the speaking rate of clear and conversational sentences; clear sentences were time-compressed to typical conversational speaking rates (200 wpm), and conversational sentences were expanded to typical clear speaking rates (100 wpm) (Picheny *et al.*, 1989; Uchanski *et al.*, 1996; Liu and Zeng, 2006). Even after accounting for processing artifacts, none of the time-scaling procedures produced speech that was more intelligible than unprocessed conversational speech (although nonuniform time-scaling, which altered the duration of phonetic segments based on segmental-level durational differences between conversational and clear speech, was less harmful to intelligibility than uniform time-scaling). Using the same systematic approach, other studies have evaluated the role of pauses in the intelligibility advantage provided by clear speech. Results suggest that introducing pauses to conversational speech does not improve intelligibility substantially, unless the short-term signal-to-noise ratio (SNR) is increased as a result of the manipulation (Uchanski *et al.*, 1996; Liu and Zeng, 2006).

From these artificial manipulations of clear and conversational speech, it can thus be concluded that neither speaking rate nor pause structure alone is responsible for a large portion of the intelligibility benefit provided by clear speech. In fact, results obtained for naturally produced clear speech are consistent with this conclusion. With training, talkers can learn to produce a form of clear speech at normal speaking rates (Krause and Braid, 2002). This speaking style, known as *clear/normal* speech, is comparable to conversational speech in both speaking rate and pause structure (Krause and Braid, 2004). Nonetheless, it provides normal-hearing listeners (in noise) with 78% of the intelligibility benefit afforded by typical clear speech (i.e., 14 percentage points for clear/normal speech vs 18 percentage points for clear speech produced at slower speaking rates; Krause and Braid, 2002).

Taken together, these results suggest that clear speech has some inherent acoustic properties, independent of rate, that account for a large portion of its intelligibility advantage. In addition, the advent of clear/normal speech has simplified the task of isolating these properties, since conversational and clear speech produced at the same speaking rate can be compared directly. Such comparisons have revealed a number of acoustical differences between clear/normal speech and conversational speech that may be associated with the differences in intelligibility between the two speaking styles (Krause and Braid, 2004). In particular, the two properties of clear/normal speech that appear most likely responsible for its intelligibility advantage are increased en-

ergy in the 1000–3000-Hz range of long-term spectra and increased modulation depth of low-frequency modulations of the intensity envelope (Krause and Braid, 2004). In an attempt to determine the extent to which these properties influence intelligibility, the present study examines the effects of two signal processing transformations designed to manipulate each of these characteristics of conversational speech independently. One transformation alters spectral characteristics of conversational speech by raising formant amplitudes typically found between 1000 and 3000 Hz, while the other alters envelope characteristics by increasing the modulation depth of low frequencies (<3–4 Hz) in several of the octave-band intensity envelopes.

In this paper, each of these signal processing transformations is described, verified, and independently applied to conversational speech. Intelligibility results are then reported for both types of processed speech as well as naturally produced conversational and clear/normal speech. The goal of the intelligibility tests was to determine the relative contribution of each of these properties to the intelligibility advantage of clear speech. If altering either of these properties improved intelligibility substantially without altering speaking rate, the corresponding transformations would provide insight into signal processing approaches for hearing aids that would have the potential to improve speech clarity as well as audibility.

II. SIGNAL TRANSFORMATIONS

Based on previously identified properties of clear/normal speech (Krause and Braid, 2004), two signal transformations were developed to alter single acoustic properties of conversational speech. Specifically, the transformations were as follows.

(1) **Transformation SPEC**— This transformation increased energy near second and third formant frequencies. These formants typically fall in the 1000–3000-Hz range (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995) where differences in long-term spectra are typically observed between conversational and clear/normal speech. The spectral differences between the two speaking styles are thought to arise from the emphasis of second and third formants in clear/normal speech; higher spectral prominences at these formant frequencies are generally evident in the short-term vowel spectra of clear/normal speech relative to conversational speech, while little spectral change in short-term consonant spectra is evident across speaking styles (Krause and Braid, 2004).

Although Transformation SPEC is similar to a high-frequency emphasis of the speech spectrum, such as what would be accomplished by frequency-gain characteristics commonly used in hearing aids, this transformation manipulates only the frequency content of vowels and other voiced segments. As a result, the increase in level of F2 and F3 relative to F1 is somewhat greater than what would result from applying a high-frequency emphasis to the entire sentence, assuming that the long-term rms level of each sentence is held constant. Because a spectral boost of this magnitude is so common in

the 1000–3000-Hz frequency range for vowels in clear/normal speech, it is important to quantify its contribution, if any, to the intelligibility advantage of clear speech.

- (2) **Transformation ENV**— This transformation increased the modulation depth of frequencies less than 3–4 Hz in the intensity envelopes of the 250-, 500-, 1000-, and 2000-Hz octave bands. This type of change is often exhibited by talkers who produce clear speech at normal rates (Krause and Braidā, 2004) and is also generally evident when talkers produce clear speech at slow rates (Payton *et al.*, 1994; Krause and Braidā, 2004; Liu *et al.*, 2004). For these reasons, and also because modulations as low as 2 Hz are known to be important for phoneme identification (Drullman *et al.*, 1994a, 1994b), the increased modulation of lower frequencies in these octave bands is considered likely to contribute to improved intelligibility (Krause and Braidā, 2004). Further evidence for this idea stems from the (speech-based) Speech Transmission Index (STI) (Houtgast and Steeneken, 1985): The speech-based STI is not only directly related to envelope spectra but also highly correlated with measured intelligibility scores for conversational speech and clear speech at a variety of speaking rates, suggesting that at least some of the differences in envelope spectra between the two speaking styles are associated with differences in intelligibility (Krause and Braidā, 2004).

Before the intelligibility effects of these transformations were measured, acoustic evaluations were conducted to verify that each transformation had produced the desired change in acoustic properties. For the purposes of these acoustic evaluations, each transformation was applied to the conversational speech of the two talkers (T4 and T5) analyzed in Krause and Braidā (2004) so that the acoustic properties of the processed speech could be directly compared to previously reported acoustic data for the clear/normal speech of these same two talkers (Krause and Braidā, 2004). In that study, speech was drawn from a corpus of nonsense (grammatically correct but semantically anomalous) sentences previously described by Picheny *et al.* (1985), and one set of 50 nonsense sentences per talker was analyzed in both clear/normal and conversational speaking styles, such that 200 utterances (100 unique sentences) were analyzed between the two talkers. For the acoustic evaluations in this study, each transformation was applied to the 100 conversational utterances analyzed in that study. Thus, for each of the two talkers, it was possible to compare the acoustic properties of the 50 processed sentences to the acoustic properties previously reported for the exact same 50 sentences spoken in both conversational and clear/normal speaking modes.

A. Transformation SPEC: Formant frequencies

The first processing scheme, Transformation SPEC, increased energy near the second and third formants by first modifying the magnitude of the short-time Fourier transform (STFT) and then using the Griffin–Lim (Griffin and Lim, 1984) algorithm to estimate a signal from its modified STFT magnitude. The STFT magnitude, or spectrogram, was com-

puted using an 8-ms Hanning window with 6 ms of overlap. The formant frequencies were then measured at 10-ms intervals for voiced portions of the speech signal using the formant tracking program provided in the ESPS/WAVES+ software package. For each 10-ms interval where voicing was present, the spectrogram magnitude was multiplied by a modified Hanning window, $w[F]$, whose endpoints in frequency, F_{start} and F_{end} , were calculated as follows:

$$F_{\text{start}} = F_2 - \min\left(2\text{BW}_2, \frac{F_1 + F_2}{2}\right), \quad (1)$$

$$F_{\text{end}} = F_3 + \min\left(2\text{BW}_3, \frac{F_2 + F_3}{2}\right), \quad (2)$$

where F_2 , F_3 , BW_2 , and BW_3 are the second and third formants and their bandwidths, respectively. A Hanning window spanning this frequency range, $h[F]$, was modified according to the following formula:

$$w[F] = Ah[F] + 1, \quad (3)$$

where A , the scale factor used to control the amount of amplification, was set to 2 in order to achieve an energy increase comparable in magnitude to that previously reported between conversational and clear/normal speech (Krause and Braidā, 2004). Finally, the Griffin–Lim (Griffin and Lim, 1984) iterative algorithm was used to derive the processed speech signal from the modified spectrogram, and the resulting sentences were then normalized for long-term rms value.

In order to evaluate whether Transformation SPEC achieved the desired effect on energy in the 1000–3000-Hz range, the long-term (sentence-level) and short-term (phonetic-level) spectra of processed speech were then computed for the two talkers from Krause and Braidā (2004) and compared to the spectra previously obtained for these talkers' conversational and clear/normal speech (Krause and Braidā, 2004). As in Krause and Braidā (2004), FFTs were computed for each windowed segment (25.6 ms non-overlapping Hanning windows) within a sentence, and then the rms average magnitude was determined over 50 sentences. A 1/3-octave representation of the spectra was obtained by summing components over 1/3-octave intervals with center frequencies ranging from 62.5 to 8000 Hz.

Figure 1 shows the long-term spectral differences of clear/normal and processed modes relative to conversational speech for T4 (results for T5 were similar), demonstrating that the processing had the desired effect on the long-term spectrum. Some spectral differences are apparent between clear/normal speech and conversational speech below 1 kHz (i.e., near F_1). However, it is the spectral differences above 1 kHz that are thought to be related to the intelligibility benefit of clear/normal speech, as these differences are seen most consistently across talkers and vowel contexts (Krause and Braidā, 2004). In this frequency range, the processed speech exhibits roughly the same increase in energy, relative to conversational/normal speech, as clear/normal speech does. Inspection of short-term spectra confirmed that this ef-

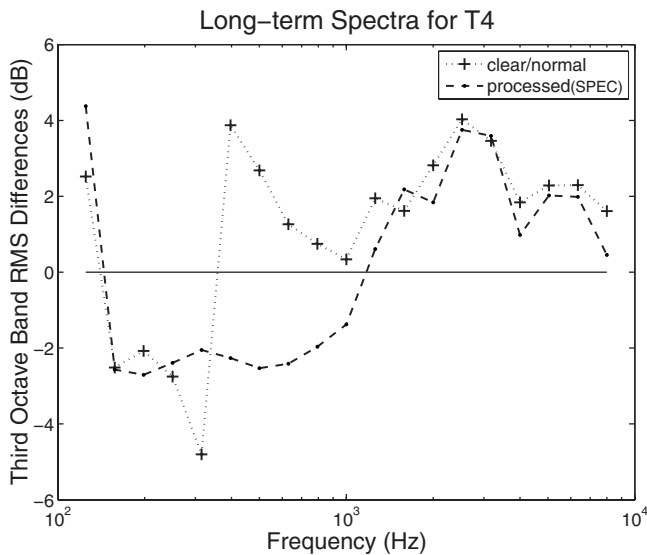


FIG. 1. Third-octave band long-term rms spectral differences for T4, a talker whose clear/normal speech has been previously analyzed in Krause and Braidá (2004). Spectral differences were obtained by subtracting the conversational (i.e., unprocessed) spectrum from the clear/normal and processed spectra.

fect was due to the expected spectral changes near the second and third formants of vowels, with little change in consonant spectra.

B. Transformation ENV: Temporal envelope

Transformation ENV was designed to increase the modulation depth for frequencies less than 3–4 Hz in the intensity envelope of the 250-, 500-, 1000-, and 2000-Hz octave bands using an analysis-synthesis approach (Drullman *et al.*, 1994b). In the analysis stage, as in Krause and Braidá (2004), 50 sentences from a talker were first concatenated and filtered into seven component signals, using a bank of fourth-order octave-bandwidth Butterworth filters, with center frequencies of 125–8000 Hz. The filters used in this stage of processing were designed so that the overall response of the combined filters was roughly ± 2 dB over the entire range of speech frequencies. The filter bank outputs for each of the seven octave bands were then squared and low-pass filtered by an eighth-order Butterworth filter with a 60-Hz cutoff frequency in order to obtain relatively smooth intensity envelopes. After downsampling the intensity envelopes in each octave band by a factor of 100, a 1/3-octave representation of the power spectra for each band was computed, with center frequencies ranging from 0.4 to 20 Hz. Finally, the power spectra were normalized by the mean of the envelope function (Houtgast and Steeneken, 1985), such that a single 100% modulated sine-wave would result in a modulation index of 1.0 for the 1/3-octave band corresponding to the modulation frequency and 0.0 for the other 1/3-octave bands.

In the second stage of processing, the envelope of each sentence was processed separately. For the octave bands in which modification was desired, the original envelope of each sentence was modified by a 200-point FIR filter designed to amplify frequencies between 0.5 and 4 Hz. The

amount of amplification was set so that the resulting modulation depths for these frequencies would be at least as large as those found in clear/normal speech and would generally fall within the range of values previously reported for clear speech, regardless of speaking rate (Krause and Braidá, 2004). This range of modulation depths was targeted because the envelope spectra of both clear/normal speech and clear speech produced at slow rates (i.e., clear/slow speech) are associated with improved intelligibility (Krause and Braidá, 2002). After the filter was applied, the modified envelope was adjusted to have the same average intensity as the original sentence envelope, and then any negative values of the adjusted envelope were set to zero. If resetting the negative portions of the envelope to zero affected the average intensity substantially, the intensity adjustment procedure was repeated until the average intensity of the modified envelope was within 0.5% of the average intensity of the original envelope. The modified envelope and original envelope were then upsampled to the original sampling rate of the signal in order to prepare for the final synthesis stage of processing.

Although it was convenient to work with intensity envelopes in the first two stages of processing so that the desired intensity envelope spectra could be achieved in the modified signal, the *amplitude* envelope was necessary for synthesis. Therefore, during the final processing stage, the time-varying ratio of the amplitude envelopes was calculated by comparing the square-root of the modified intensity envelope with the square-root of the original intensity envelope. The original octave-band signals were then transformed by multiplying the original signal in each octave band (with fine structure) by the corresponding time-varying amplitude ratio for that band. In order to ensure that no energy outside the octave band was inadvertently amplified, the result was also low-pass filtered by a fourth-order Butterworth filter with cutoff frequency corresponding to the upper cutoff frequency for the octave band. The processed version of the signal was then obtained by summing the signals in each octave band. Lastly, the processed sentences were normalized for long-term rms level.

After synthesis was completed, it was determined through informal listening tests that these modifications caused the speech of the female talker (T4) to sound more unnatural than the male talker (T5). The transformation was applied to two additional female talkers with the same results. A likely explanation for this problem was thought to be that the fundamental frequency of the female talkers tends to fall in the second octave band, and amplifying slowly varying modulations of voicing is not likely to occur in natural speech unless the talker slows down. This explanation was supported by the acoustic data, since an increase in modulation index was not exhibited in the 250-Hz band for clear/normal speech for the female talker, T4, although it was present in the 500-Hz band (Krause and Braidá, 2004). Informal listening tests confirmed that eliminating the envelope modification in the 250-Hz band improved the quality of the female talkers' processed speech. Therefore, the signal transformation procedure was specified to modify only the 500-, 1000-, and 2000-Hz bands for female talkers.

In order to evaluate the effect of the processing on the

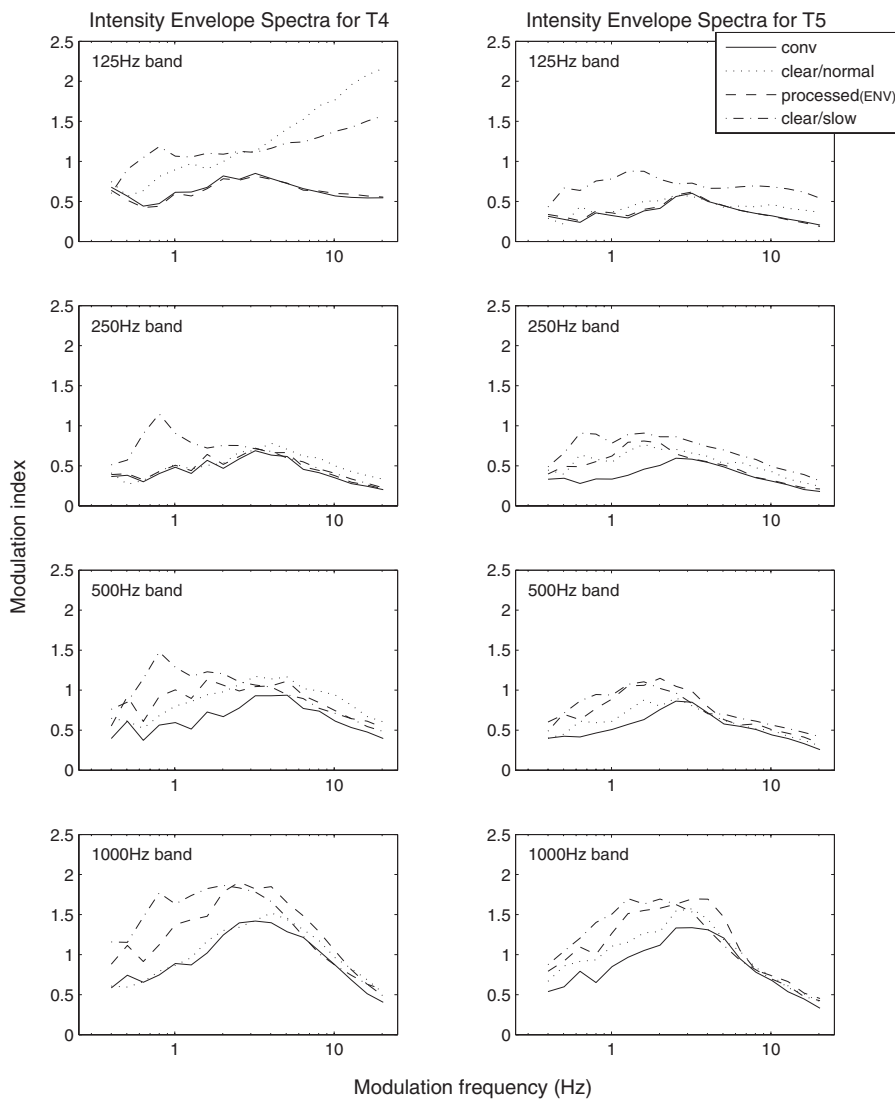


FIG. 2. Spectra of intensity envelopes in the lower four octave bands for T4 and T5, talkers whose clear/normal and clear/slow speech have been previously analyzed in Krause and Braida (2004).

intensity envelopes, the envelope spectra of the processed speech were computed for the two talkers from Krause and Braida (2004) and compared to the envelope spectra previously obtained for these talkers' unprocessed (conversational) speech, as well as their clear speech at both normal and slow rates (clear/normal and clear/slow). The spectra of the octave-band intensity envelopes for both talkers in each speaking style are shown in Figs. 2 and 3. From these figures, it can be seen that the processing had the intended effect on the spectra of the octave-band intensity envelopes, with envelope spectra of processed speech falling between that of previously measured clear/normal and clear/slow spectra (Krause and Braida, 2004) for frequencies less than 3–4 Hz in the specified octave bands (500-, 1000-, and 2000 Hz for both talkers as well as 250 Hz for T5, the male talker) and no substantial changes in the remaining octave bands.

III. INTELLIGIBILITY TESTS

To assess the effectiveness of the signal transformations, the intelligibility of the processed speech was measured and compared to the intelligibility of naturally produced conversational speech and clear speech at normal rates. Intelligibil-

ity in each speaking style was measured by presenting processed and unprocessed speech stimuli to normal-hearing listeners in the presence of wideband noise as well as to hearing-impaired listeners in a quiet background.

A. Speech stimuli

Speech stimuli were generated from speech materials recorded for an earlier study of clear speech elicited naturally at normal speaking rates (Krause and Braida, 2002). In that study, nonsense sentences (e.g., *His right cane could guard an edge.*) from the Picheny corpus (Picheny *et al.*, 1985) were recorded by five talkers in a variety of speaking styles. Materials were selected from one male (T5) and three female (T1, T3, and T4) talkers, because these four talkers obtained relatively large intelligibility benefits from clear/normal speech (11–32 percentage points relative to conversational speech) and were therefore most likely to benefit from signal transformations based on its acoustic properties. The materials selected for T4 and T5 were generally different than those used for these talkers in the acoustic evaluations of the transformations, except as noted below.

For each of the four talkers, 90 sentences recorded in a conversational speaking style and 30 sentences recorded in a

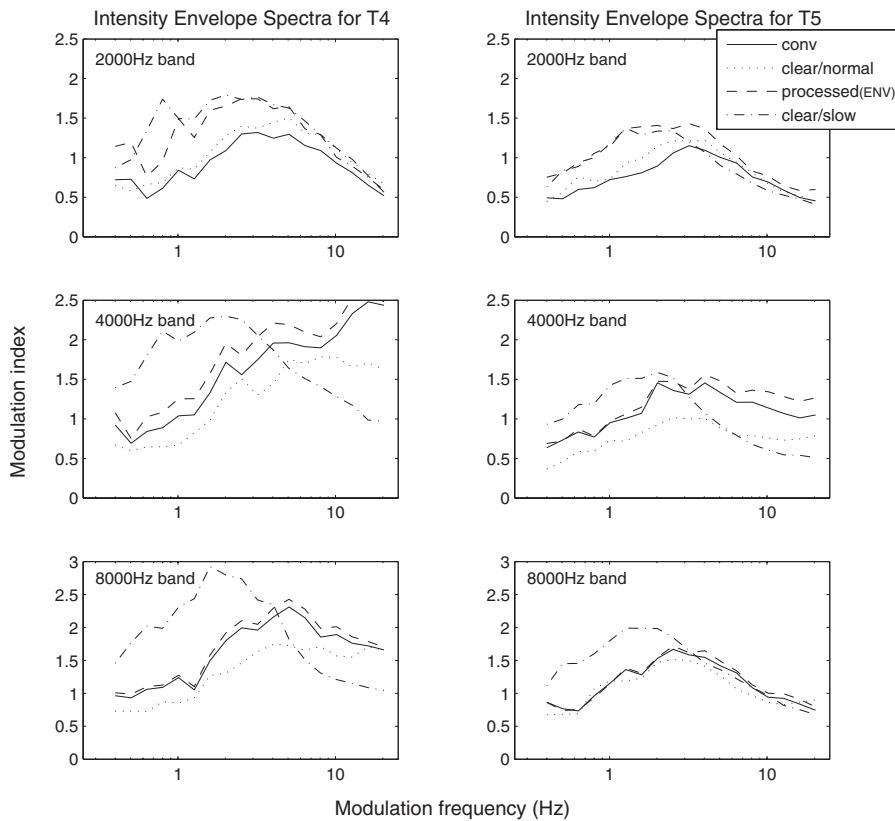


FIG. 3. Spectra of intensity envelopes in the upper three octave bands for T4 and T5, talkers whose clear/normal and clear/slow speech have been previously analyzed in Krause and Braida (2004).

clear/normal speaking style were used for this experiment. In one case (T1), 30 clear sentences that had been elicited at a quick speaking rate were used rather than those elicited in the clear/normal style, since these sentences were produced at about the same speaking rate as the clear/normal sentences but received higher intelligibility scores (Krause and Braida, 2002). Of the 90 conversational sentences used for each talker, 30 remained unprocessed, 30 were processed by Transformation SPEC, and 30 were processed by Transformation ENV (for T4 and T5, 10 of these sentences had been used previously in the acoustical evaluations), resulting in 30 unique sentences per condition per talker. Because different sentences were used for each talker, a total of 480 unique sentences (30 sentences \times 4 conditions \times 4 talkers) were thus divided evenly between the four different speaking conditions: conversational, processed(SPEC), processed(ENV), and clear/normal. The two conditions that were naturally produced (i.e., unprocessed) were included as reference points for the processed conditions, with conversational speech representing typical intelligibility and clear/normal

speech representing the maximum intelligibility that talkers can obtain naturally by speaking clearly without altering speaking rate.

B. Listeners

Eight listeners were recruited from the MIT community to evaluate the intelligibility of the speech stimuli. All of the listeners were native speakers of English who possessed at least a high school education. Five of the listeners (one male and four females; age range: 19–43 years) had normal hearing, with thresholds no higher than 20 dB HL for frequencies between 250 and 4000 Hz, while three of the listeners (three males; age range: 40–65 years) had stable sensorineural hearing losses that were bilateral and symmetric. The audiometric characteristics for the test ears of the hearing-impaired listeners are summarized in Table I. For these listeners, the intelligibility tests were administered either to the ear with

TABLE I. Audiometric characteristics for the test ears of the hearing-impaired listeners. NR indicates no response to the specified frequency.

Listener	Sex	Age	Test ear	Word ^a recognition (%)	Thresholds (dB HL)					
					250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
L6HI	M	65	Left	100	55	60	45	45	55	85
L7HI	M	64	Left	92	10	20	40	60	65	NR
L8HI	M	40	Right	100	50	55	55	60	90	85

^aEither NU-6 (Tillman and Carhart, 1996) or W-22 (Hirsh et al., 1952).

better word recognition performance during audiometric testing or to the preferred ear if no such difference in word recognition performance was observed.

It is worth noting here that the purpose of the intelligibility tests was not to make comparisons between hearing-impaired listeners and normal-hearing listeners but simply to have all eight listeners evaluate the intelligibility benefit of clear/normal speech and the two signal-processing transformations. For nonsense sentences such as those used in this study, it has been shown that the intelligibility *benefit* of clear speech (relative to conversational speech) is roughly the same for normal-hearing listeners (in noise) as for hearing-impaired listeners (in quiet), despite differences in absolute performance levels (Uchanski *et al.*, 1996). For example, Uchanski *et al.* (1996) reported that clear speech improved intelligibility by 15–16 points on average for both normal-hearing listeners (clear: 60% vs conversational: 44%) and hearing-impaired listeners (clear: 87% vs conversational: 72%), including listeners with audiometric configurations similar to the listeners in this study. Similarly, Payton *et al.* (1994) reported that the clear speech intelligibility benefit obtained by each of their two hearing-impaired listeners fell within the range of benefits obtained by the ten normal-hearing listeners in that study. In both cases, the percentage change in intelligibility was different for the two groups [larger for normal-hearing listeners in the study by Uchanski *et al.* (1996); larger for hearing-impaired listeners in the study by Payton *et al.* (1994)], but the intelligibility benefit (absolute difference in percentage points between clear and conversational speech) was roughly the same. Given these data, listeners were not divided on the basis of hearing status in this study, because the only intelligibility measures planned were measures of intelligibility benefit.

C. Procedures

All listeners were tested monaurally over TDH-39 headphones. Normal-hearing listeners were tested in the presence of wideband noise, and hearing-impaired listeners were tested in quiet. As described above, clear speech typically provides about the same amount of benefit to both types of listeners under these test conditions (e.g., Payton *et al.*, 1994; Uchanski *et al.*, 1996).

Normal-hearing listeners were seated together in a sound-treated room and tested simultaneously. Each normal-hearing listener selected the ear that would receive the stimuli and was encouraged to switch the stimulus to the other ear when fatigued. For stimulus presentation, stereo signals were created for each sentence, with speech on one channel and speech-shaped noise (Nilsson *et al.*, 1994) of the same rms level on the other channel. The speech was attenuated by 1.8 dB and added to the speech-shaped noise, and the resulting signal (SNR = -1.8 dB) was presented to the listeners from a PC through a Digital Audio Labs (DAL) soundcard.

Hearing-impaired listeners were tested individually in a sound-treated room. A linear frequency-gain characteristic was obtained for each hearing-impaired listener using the NAL-R procedure (Byrne and Dillon, 1986) and then imple-

mented using a third-octave filter bank (General Radio, model 1925). This procedure provided frequency-shaping and amplification based on the characteristics of the individual's hearing loss. At the beginning of each condition, the listener was also given the opportunity to adjust the overall system gain. The speech was then presented through the system to the listener from a DAL card on a PC. These procedures ensured that the presentation level of each condition was both comfortable and as audible as possible for each hearing-impaired listener.

Because it has been shown that learning effects for these materials are minimal (Picheny *et al.*, 1985), all listeners heard the same conditions in the same order. However, the presentation order for test conditions was varied across talkers such that no condition was consistently presented first (or last). Listeners were presented a total of sixteen 30-sentence lists (4 talkers \times 4 conditions/talker) and responded by writing their answers on paper. They were given as much time as needed to respond but were presented each sentence only once. Intelligibility scores were based on the percentage of key words (nouns, verbs, and adjectives) identified correctly, using the scoring rules described by Picheny *et al.* (1985).

IV. RESULTS

The clear/normal speaking condition was most intelligible overall at 58% key words correct, providing a 13 percentage point improvement in intelligibility over conversational speech (45%). The size of this intelligibility advantage was consistent with the 14 percentage point advantage of clear/normal speech measured previously (Krause and Braid, 2002) for normal-hearing listeners in noise. Neither of the signal transformations, however, provided nearly as large of an intelligibility benefit as clear/normal speech. At 49%, processed(SPEC) speech was just 14 points more intelligible than conversational speech on average, while processed(ENV) speech (24%) was considerably *less* intelligible than conversational speech.

An analysis of variance performed on key word scores (after an arcsine transformation to equalize variances) showed that the main effect of condition was significant [$F(3,256)=273$, $p < 0.01$] and accounted for the largest portion of the variance ($\eta^2=0.371$) in intelligibility. Post-hoc tests with Bonferroni corrections confirmed that overall differences between all conditions were significant at the 0.05 level. As expected, the main effects of listener [$F(7,256)=90$, $p < 0.01$] and talker [$F(3,256)=62$, $p < 0.01$] were also significant (since, in general, some talkers will be more intelligible than others and some listeners will perform better on intelligibility tasks than others), but the listener \times talker interaction was not significant. For the purposes of this study, it is more important to note that all interactions of talker and listener with condition were significant but accounted for relatively small portions of the variance (among these terms, listener \times talker \times condition accounted for the largest portion of the variance, with $\eta^2=0.057$). Nonetheless, examination of these interactions provides additional insight regarding the relative effectiveness of each signal transformation.

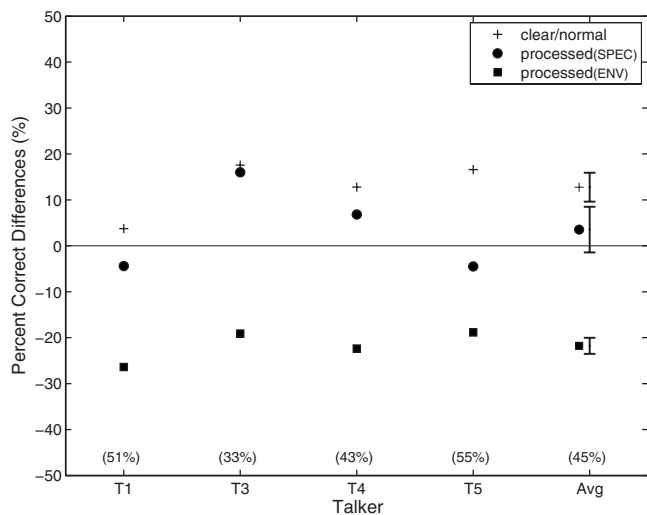


FIG. 4. Percent correct difference scores for each speaking condition (relative to conversational speech) obtained by individual talkers, averaged across listeners. Baseline conversational intelligibility scores are listed in parentheses. Errorbars at right indicate standard error of talker difference scores in each condition.

Figure 4 shows that the effects of condition were largely consistent across individual talkers. For example, all talkers achieved a benefit with clear/normal speech, and the benefit was generally sizable (>10 points for three of the four talkers; note that this result was expected since the talkers were selected because they had previously demonstrated intelligibility improvements with clear/normal speech). Similarly, Transformation ENV was detrimental to intelligibility for all talkers: processed(ENV) speech was substantially less intelligible (19–27 points) than conversational speech. For Transformation SPEC, however, the benefit was not consistent across talkers. Instead, processed(SPEC) speech showed improved intelligibility for two talkers (T3: 16 points; T4: 7 points) and reduced intelligibility (relative to conversational speech) for the other two talkers (T1: -5 points; T5: -5 points).

Figure 5 shows the effects of condition across individual listeners. Again, the effects of clear/normal speech and processed(ENV) speech were generally consistent. For clear/normal speech, seven of eight listeners received an intelligibility benefit, with the average benefit across talkers ranging from 8 to 19 percentage points. The effect of Transformation ENV on listeners was similarly robust, although in the opposite direction. Processed(ENV) speech was less intelligible than conversational speech for all listeners, with differences in intelligibility ranging from -11 to -41 percentage points on average across talkers. The effect of Transformation SPEC, on the other hand, was less consistent. Although processed(SPEC) speech was more intelligible than conversational speech on average for six of eight listeners, these listeners did not receive a benefit from the processed(SPEC) speech of all talkers. All listeners received large benefits from processed(SPEC) speech for T3 (5–25 points), but none received any benefit for T1, and only about half received benefits of any size for T4 and T5, respectively. As a result, Fig. 5 shows that the size of the average intelligibility benefit

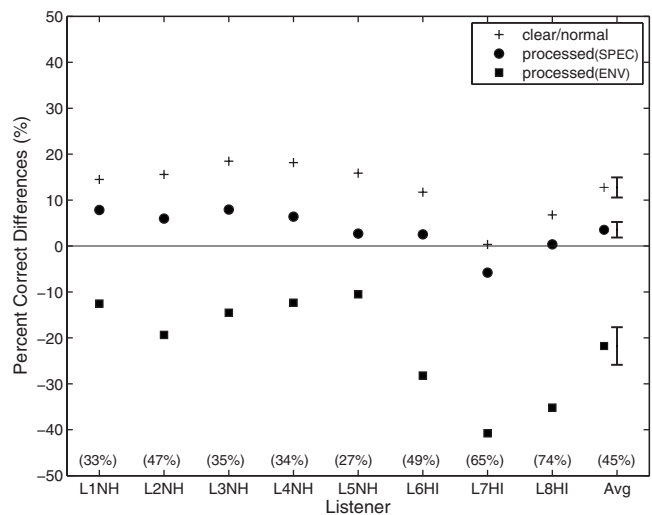


FIG. 5. Percent correct difference scores for each speaking condition (relative to conversational speech) obtained by individual listeners (NH designates normal-hearing listeners and HI designates hearing-impaired listeners), averaged across talkers. Baseline conversational intelligibility scores are listed in parentheses. Errorbars at right indicate standard error of listener difference scores in each condition.

that each listener received from processed(SPEC) speech was considerably smaller (3–8 points) than the benefit received from clear/normal speech.

With the exception of the processed(SPEC) speech of T4 and T5, the relative intelligibility of speaking conditions for any given talker was qualitatively very similar across individual listeners. For example, the benefit of clear/normal speech was so robust that the seven listeners who received a benefit from clear/normal speech on average across talkers (Fig. 5) also received an intelligibility benefit from each individual talker, and the remaining listener (L7HI) received a comparable benefit for two of the four talkers (T3 and T4) as well. Thus, clear/normal speech improved intelligibility for nearly all (30) of the 32 combinations of individual talkers and listeners, in most cases (24) by a substantial margin (>5 percentage points). Similarly, individual data confirm that Transformation ENV consistently decreased intelligibility; with only one exception (L5NH for T5), processed(ENV) speech was substantially less intelligible (<-5 percentage points) than conversational speech for each of the 32 combinations of individual talkers and listeners. Such consistency across four talkers and eight listeners (particularly listeners who differ considerably in age and audiometric profile) clearly shows that (1) Transformation ENV is detrimental to intelligibility, and (2) the benefit from clear/normal speech is considerably larger and more robust than the benefit of Transformation SPEC.

A. Hearing-impaired listeners

Although the relative intelligibility of conditions was qualitatively similar across listeners in general, a few differences were observed between hearing-impaired and normal-hearing listeners. In particular, the benefit of clear/normal speech was smaller for hearing-impaired listeners on average (7 points) than for normal-hearing listeners (17 points), and the detrimental effect of Transformation ENV was larger

(-35 vs -14 points). In addition, all three hearing-impaired listeners exhibited substantial decreases in intelligibility for the processed(SPEC) speech of T1 (-7 points on average) and T5 (-19 points) that were not typical of the normal-hearing listeners, who merely exhibited little to no benefit (T1: -3 points; T5: 4 points). As a result, Transformation SPEC did not improve intelligibility on average for hearing-impaired listeners but did provide a 6-point improvement for normal-hearing listeners, even though the processing provided both groups with comparable intelligibility improvements for T3 (17 points for NH listeners vs 14 points for HI listeners) and T4 (6 points vs 8 points). Whether any of these trends could reflect true difference(s) between populations cannot be determined from this study, because listeners were recruited without regard to audiometric profile. While these listeners provided valuable information regarding the relative intelligibility of the two signal transformations that were developed, more listeners (both hearing-impaired and normal-hearing) would be required to detect differences in performance between groups of listeners with different audiometric profiles.

B. Processing artifacts

To assess whether the potential benefits of either signal transformation may have been reduced or obscured by digital-signal processing artifacts, additional listeners were employed to evaluate speech that had been processed twice. For each signal transformation, twice-processed speech was obtained by first altering the specified acoustic parameter and then restoring the parameter to its original value. Thus, any reduction in intelligibility between the original (unprocessed) speech and the restored (twice-processed) speech would reflect only those deleterious effects on intelligibility specifically caused by processing artifacts.

For each talker, five conditions were tested, one conversational and four processed conditions: processed(SPEC), processed(ENV), restored(SPEC), and restored(ENV). Four additional normal-hearing listeners (all males; age range: 21–27 years old) each heard the speech of one talker in all five conditions, and each listener was presented speech from a different talker. The presentation setup was the same as described above, with one small difference: A SNR of 0 dB was used to avoid floor effects, since scores obtained in initial intelligibility tests for processed(ENV) speech were fairly low (e.g., 14% for T3). If processing artifacts were wholly responsible for these low scores, further reductions in intelligibility would be expected for restored(ENV) speech.

Average scores for the restored(SPEC) condition (45%) were the same as average scores for unprocessed conversational speech (45%), suggesting that processing artifacts associated with Transformation SPEC were negligible. In contrast, processing artifacts were a substantial issue for Transformation ENV, as the restored(ENV) condition was 19 points less intelligible than conversational speech. Notably, processed(ENV) speech showed some benefit relative to the restored(ENV) condition for two talkers (7 points for T3 and 15 points for T5). Therefore, it may be possible to achieve intelligibility improvements by altering the temporal enve-

lope, if processing artifacts can be avoided. Also noteworthy is that while processed(SPEC) speech provided some benefit for three of four talkers in the initial experiment at SNR=-1.8 dB, the benefit was evident for only one of those talkers (T5) at SNR=0 dB. Although this difference between experiments may have occurred by chance (particularly given that only one listener per talker was used at SNR=0 dB), it is also possible that the benefit of Transformation SPEC for normal-hearing listeners diminishes as SNRs improve. This possibility will be discussed further in Sec. V.

V. DISCUSSION

Despite the fact that both signal transformations were based on the acoustics of clear/normal speech, results of intelligibility tests showed that neither transformation provided robust intelligibility improvements over unprocessed conversational speech. Transformation SPEC, which increased energy near second and third formant frequencies, improved intelligibility for some talkers and listeners, but the benefit was inconsistent and averaged just 4% overall. Transformation ENV, which enhanced low-frequency (<3–4 Hz) modulations of the intensity envelope, decreased intelligibility for all talkers and listeners, most likely because of the detrimental effects of processing artifacts associated with the transformation.

Although previous clear speech studies found intelligibility results for hearing-impaired listeners to be consistent with results for normal-hearing listeners in noise (Payton *et al.*, 1994; Uchanski *et al.*, 1996), some differences between these two populations were observed in this study. Most notably, Transformation SPEC improved intelligibility for normal-hearing listeners by 6 points on average but did not provide any benefit at all to hearing-impaired listeners on average. While both types of listeners did receive large benefits from processed(SPEC) speech relative to conversational speech for T3 (14 points for hearing-impaired listeners and 17 points for normal-hearing listeners), all three hearing-impaired listeners exhibited substantial *decreases* in intelligibility for the processed(SPEC) speech of T1 (-7 points on average) and T5 (-19 points on average) that were not typical of the normal-hearing listeners. These differences suggest that the benefit of Transformation SPEC may be associated with formant audibility for hearing-impaired listeners. That is, increasing the energy near F2 and F3 may improve the intelligibility of some talkers (e.g., T3), whose formants are not consistently audible to hearing-impaired listeners with NAL-R amplification, but may not improve the intelligibility of talkers whose formants are consistently audible (e.g., T1 and T5). Instead, listeners may perceive the processed sentences of these talkers as unnecessarily loud and respond by decreasing the overall level of amplification, thereby reducing the level of other frequency components and potentially decreasing sentence intelligibility.

In this case, it would seem likely that the benefit of formant processing for normal-hearing listeners would also be largely associated with formant audibility. To examine this possibility, band-dependent SNRs for each talker were calculated from the speech stimuli used in the intelligibility

tests. Within each of the five third-octave bands over which formant modification occurred (center frequencies ranging from 1260 to 3175 Hz), a band-dependent SNR was measured by comparing the rms level of the talker's speech within that band to the rms level of speech-shaped noise within that band. These measurements confirm that T3 had the poorest band-dependent SNRs in this frequency region (-5.3 dB on average) of all talkers, while T1 had the highest (-0.3 dB). Thus, it is not surprising that all five normal-hearing listeners received large intelligibility improvements from the processed(SPEC) speech of T3—for whom raising the level of the formants could substantially increase the percentage of formants that were audible over the noise, but none received a benefit greater than 1% from the processed(SPEC) speech of T1—for whom a high percentage of formants in conversational speech were probably already above the level of the noise. Inspection of spectrograms for each talker is consistent with this explanation: A much higher percentage of second and third formants are at levels well above the level of the noise for T1 than for T3. Following this reasoning, a higher percentage of formants would also be audible for intelligibility tests conducted at higher SNRs, which also explains why only one of four listeners obtained a benefit from processed(SPEC) speech during the follow-up intelligibility tests that examined processing artifacts, which were conducted at SNR=0 dB.

That formant audibility would play a role in the intelligibility benefit of Transformation SPEC is not surprising, given its similarities to high-frequency spectral emphasis. Such similarities also suggest that any intelligibility improvement provided by Transformation SPEC should not be large; altering the spectral slope of frequency-gain characteristics used to present sentences in noise has little effect on sentence intelligibility, even for hearing-impaired listeners (van Dijkhuizen *et al.*, 1987, 1989). Unlike Transformation SPEC, which alters only the speech signal, however, the frequency-gain characteristic affects both the speech signal and the background noise, thus preserving band-dependent SNRs. In contrast, decreased spectral tilt can occur naturally when talkers produce speech in noisy environments (Summers *et al.*, 1988), leaving background noise unchanged. Although typically associated with large improvements in intelligibility, the decreased spectral tilt occurs in this circumstance in conjunction with several other acoustic changes, similar to those observed in clear speech (Summers *et al.*, 1988). In light of the intelligibility results for Transformation SPEC, it seems likely that one or more of those acoustic changes provides the bulk of that intelligibility benefit.

While the relative intelligibility of the other conditions in this study was qualitatively similar for hearing-impaired listeners and normal-hearing listeners in noise, a second difference observed between these populations is that hearing-impaired listeners received a smaller benefit from clear/normal speech (7 vs 17 points) and a larger detriment from processed(ENV) speech (-35 vs -14 points) than their normal-hearing counterparts. Given that only three hearing-impaired listeners were tested, these differences could certainly have occurred by chance. Furthermore, there is a pos-

sibility that speaking styles were differentially affected by the NAL-R amplification provided to hearing-impaired listeners, thereby altering the relative intelligibility differences between conditions for these listeners. In other words, it is possible that the smaller benefit from clear/normal speech occurred for hearing-impaired listeners because the NAL-R amplification improved the intelligibility of conversational speech relatively more than it improved clear/normal speech. However, previous examination of these issues in clear speech produced at slower rates suggests that this possibility is not likely; that is, the benefit of clear speech is typically independent of frequency-gain characteristic (Picheny *et al.*, 1985). Therefore, further investigation regarding the benefits of clear/normal speech for hearing-impaired listeners is warranted to determine whether the trend observed in this study reflects a true difference between populations.

If such a difference between the populations exists, one possibility is that the benefits of clear/normal speech may be related to age, since the hearing-impaired listeners in this study were older (40–65 years) than the normal-hearing listeners (19–43 years). Another possibility is that an individual's audiometric characteristics may be a factor in whether clear/normal speech can be of benefit. A close inspection of interactions between hearing-impaired listener and talker reveals that L7HI received little or no benefit from clear/normal speech, except when listening to T4's speech, while each of the other two listeners experienced moderate to large intelligibility gains from clear/normal speech for all talkers. Since L7HI also had the most precipitous hearing loss, it is possible that listeners with this type of audiometric configuration may not be as likely to benefit from clear/normal speech as other hearing-impaired listeners. If so, clear/normal speech would differ in this respect from clear speech at slow rates, which provides roughly the same amount of benefit to listeners with various audiometric configurations (Uchanski *et al.*, 1996). To address the question of whether age and/or audiometric characteristics are linked to an individual's ability to benefit from clear/normal speech, additional intelligibility tests targeting younger and older groups of listeners with various configurations and severity of hearing loss would be required.

VI. CONCLUSION

Of the two processing schemes examined in this study, only Transformation SPEC, the transformation associated with modification of formant frequencies, provided an intelligibility advantage over conversational speech. However, this benefit appeared to be largely a function of formant audibility: The transformation was more likely (1) to improve intelligibility for talkers with second and third formants that were relatively low in amplitude prior to processing and (2) to provide benefits for normal-hearing listeners in noise, who could take advantage of improvements in band-dependent SNRs associated with processing. Even so, the benefit that normal-hearing listeners in noise received from processed(SPEC) speech (6 percentage points on average) was less than half what they received from clear/normal speech (17 percentage points), suggesting that increased energy near

second formant and third formants is not the only factor responsible for the intelligibility advantage of clear/normal speech for these listeners.

Another factor that is likely to contribute to the improved intelligibility of clear speech (Krause and Braid, 2004; Liu *et al.*, 2004), at least at high SNRs (Liu and Zeng, 2006), is increased depth of low-frequency modulations of the intensity envelope. Although Transformation ENV successfully increased the depth of these modulations in the speech signal, processing artifacts made it difficult to determine the extent to which, if at all, this factor accounts for the clear speech advantage. While processed(ENV) speech was less intelligible than unprocessed conversational speech, it was more intelligible than the restored(ENV) condition for two of four talkers, suggesting that intelligibility improvements associated with altering the temporal intensity envelope may be possible, if processing artifacts can be minimized. Given that Transformation ENV manipulated all low-frequency modulations uniformly, it is also possible that an unnatural prosodic structure imposed by the transformation is the source of the processing artifacts. If so, it may be helpful to develop a processing tool that allows for nonuniform alteration of the intensity envelope. With such a tool, it may be possible to enhance low-frequency modulations while maintaining the general prosodic structure of the speech. Speech manipulated in this manner could then be evaluated with further intelligibility tests in order to provide a better understanding of how increases in intensity envelope modulations are related to the clear speech benefit.

Although further research is needed, the results of the present study are an essential first step toward quantifying the role of spectral and envelope characteristics in the intelligibility advantage of clear speech. By independently manipulating these acoustic parameters and systematically evaluating the corresponding effects on intelligibility, two important findings have been established. First, an increase in energy between 1000 and 3000 Hz does not fully account for the intelligibility benefit of clear/normal speech. One or more other acoustic factors must also play a role. Second, simple filtering of the intensity envelope to achieve increased depth of modulation is generally detrimental to intelligibility, even though this acoustic property is considered likely to be at least partly responsible for the intelligibility benefit of clear speech (Krause and Braid, 2004; Liu *et al.*, 2004). Therefore, future research investigating nonuniform alterations of the intensity envelope is required to isolate the effects of this factor on intelligibility. In addition, signal transformations and intelligibility tests aimed at identifying the role of other acoustic properties of clear/normal speech (Krause and Braid, 2004) are also needed. Such tests would provide additional information regarding the mechanisms responsible for the intelligibility benefit of clear speech and could ultimately lead to improved signal processing approaches for digital hearing aids.

ACKNOWLEDGMENTS

The authors wish to thank Jae S. Lim, Joseph S. Perkell, Kenneth N. Stevens, and Rosalie M. Uchanski for many

helpful technical discussions. In addition, they thank J. Desloge for his assistance in scoring sentences. Financial support for this work was provided by a grant from the National Institute on Deafness and Other Communication Disorders (NIH Grant No. 5 R01 DC 00117).

- Bradlow, A. R., and Bent, T. (2002). "The clear speech effect for non-native listeners," *J. Acoust. Soc. Am.* **112**, 272–284.
- Bradlow, A. R., Kraus, N., and Hayes, E. (2003). "Speaking clearly for children with learning disabilities: Sentences perception in noise," *J. Speech Lang. Hear. Res.* **46**, 80–97.
- Byrne, D., and Dillon, H. (1986). "The national acoustic laboratories new procedure for selecting the gain and frequency response of a hearing aid," *Ear Hear.* **7**, 257–265.
- Drullman, R., Festen, J. M., and Plomp, R. (1994a). "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.* **95**, 2670–2680.
- Drullman, R., Festen, J. M., and Plomp, R. (1994b). "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.* **95**, 1053–1064.
- Ferguson, S. H., and Kewley-Port, D. (2002). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **112**, 259–271.
- Griffin, D. W., and Lim, J. S. (1984). "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.* **32**, 236–243.
- Helfer, K. S. (1998). "Auditory and auditory-visual recognition of clear and conversational speech by older adults," *J. Am. Acad. Audiol.* **9**, 234–242.
- Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (1995). "Acoustic characteristics of american english vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Eldert, E., and Benson, R. W. (1952). "Development of materials for speech audiometry," *J. Speech Hear. Disord.* **17**, 321–337.
- Houtgast, T., and Steeneken, H. (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.
- Krause, J. C., and Braid, L. D. (2002). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," *J. Acoust. Soc. Am.* **112**, 2165–2172.
- Krause, J. C., and Braid, L. D. (2003). "Effects of listening environment on intelligibility of clear speech at normal speaking rate," *Iran. Audiol.* **2**, 39–47.
- Krause, J. C., and Braid, L. D. (2004). "Acoustic properties of naturally produced clear speech at normal speaking rates," *J. Acoust. Soc. Am.* **115**, 362–378.
- Liu, S., and Zeng, F.-G. (2006). "Temporal properties in clear speech perception," *J. Acoust. Soc. Am.* **120**, 424–432.
- Liu, S., Rio, E. D., and Zeng, F.-G. (2004). "Clear speech perception in acoustic and electric hearing," *J. Acoust. Soc. Am.* **116**, 2374–2383.
- Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.* **95**, 1085–1099.
- Payton, K. L., Uchanski, R. M., and Braid, L. D. (1994). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* **95**, 1581–1592.
- Peterson, G., and Barney, H. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Picheny, M. A., Durlach, N. I., and Braid, L. D. (1985). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," *J. Speech Hear. Res.* **28**, 96–103.
- Picheny, M. A., Durlach, N. I., and Braid, L. D. (1986). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," *J. Speech Hear. Res.* **29**, 434–446.
- Picheny, M. A., Durlach, N. I., and Braid, L. D. (1989). "Speaking clearly for the hard of hearing III: An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech," *J. Speech Hear. Res.* **32**, 600–603.
- Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (1988). "Effects of noise on speech production: Acoustic and perceptual analyses," *J. Acoust. Soc. Am.* **84**, 917–928.
- Tillman, T., and Carhart, R. (1966). "An expanded test for speech discrimination utilizing cnc monosyllabic words: Northwestern university auditory

test no. 6," Technical Report No. SAM-TR-66-55, USAF School of Aerospace Medicine, Brooks Air Force Base, TX.

Uchanski, R. M., Choi, S., Braida, L. D., Reed, C. M., and Durlach, N. I. (1996). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," *J. Speech Hear. Res.* **39**, 494–509.

van Dijkhuizen, J., Anema, P., and Plomp, R. (1987). "The effect of varying

the slope of the amplitude-frequency response on the masked speech-reception threshold of sentences," *J. Acoust. Soc. Am.* **81**, 465–469.

van Dijkhuizen, J. N., Festen, J. M., and Plomp, R. (1989). "The effect of varying the slope of the amplitude-frequency response on the masked speech-reception threshold of sentences for hearing-impaired listeners," *J. Acoust. Soc. Am.* **86**, 621–628.