

The effects of selective consonant amplification on sentence recognition in noise by hearing-impaired listeners

Rithika Saripella and Philipos C. Loizou^{a)}

Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75080

Linda Thibodeau and Jennifer A. Alford

Callier Center for Communication Disorders, Dallas, Texas 75235

(Received 21 November 2010; revised 22 August 2011; accepted 25 August 2011)

Weak consonants (e.g., stops) are more susceptible to noise than vowels, owing partially to their lower intensity. This raises the question whether hearing-impaired (HI) listeners are able to perceive (and utilize effectively) the high-frequency cues present in consonants. To answer this question, HI listeners were presented with clean (noise absent) weak consonants in otherwise noise-corrupted sentences. Results indicated that HI listeners received significant benefit in intelligibility (4 dB decrease in speech reception threshold) when they had access to clean consonant information. At extremely low signal-to-noise ratio (SNR) levels, however, HI listeners received only 64% of the benefit obtained by normal-hearing listeners. This lack of equitable benefit was investigated in Experiment 2 by testing the hypothesis that the high-frequency cues present in consonants were not audible to HI listeners. This was tested by selectively amplifying the noisy consonants while leaving the noisy sonorant sounds (e.g., vowels) unaltered. Listening tests indicated small (~10%), but statistically significant, improvements in intelligibility at low SNR conditions when the consonants were amplified in the high-frequency region. Selective consonant amplification provided reliable low-frequency acoustic landmarks that in turn facilitated a better lexical segmentation of the speech stream and contributed to the small improvement in intelligibility. © 2011 Acoustical Society of America. [DOI: 10.1121/1.3641407]

PACS number(s): 43.71.Ky, 43.71.Es [EB]

Pages: 3028–3037

I. INTRODUCTION

Vowel perception generally poses little difficulty for hearing-impaired listeners due partially to the fact that the level of the vowels is much greater than the level of the consonants (Owens *et al.*, 1968; Edwards, 2004). In contrast, consonant perception is considerably much more challenging for hearing-impaired listeners (e.g., Owens *et al.*, 1972). Among other factors, consonant perception by hearing-impaired (HI) listeners seems to be affected by multiband compression (Yund and Buckles, 1995; Edwards, 2004), which reduces spectral contrast, and the degree of hearing loss, particularly in the high-frequency regions where some consonants (e.g., /s/, /t/) have prominent energy. This reduced spectral contrast combined with the hearing loss clearly influences the audibility of consonants (Owens *et al.*, 1972; Turner and Robb, 1987). Simple amplification of consonants, however, to restore audibility may not always lead to intelligibility benefit for HI listeners. As hypothesized by many (e.g., Skinner, 1980; Hogan and Turner, 1998; Turner and Cummings, 1999; Ching *et al.*, 1998; Moore, 2001), once the hearing loss in a particular region of the cochlea becomes too severe (beyond 55 dB HL, according to some studies), speech information is affected by distortion, even when presented at suprathreshold levels.

The intensity of some consonants can be as low as 20 dB lower than that of vowels (Gordon-Salant, 1986;

Freyman *et al.*, 1991). For that reason, a number of studies has considered selective amplification of consonants, while leaving the vowel level constant, and has examined the role of the consonant-vowel intensity ratio (CVR) on consonant identification. Increasing the CVR has been found to improve consonant recognition performance in normal-hearing (NH) listeners (Gordon-Salant, 1986; Freyman and Nerbonne, 1989) and in patients with sensorineural hearing loss (Gordon-Salant, 1987; Montgomery and Edge, 1988). In conditions wherein listeners were forced to rely more on temporal-envelope cues rather than on spectral cues, Freyman *et al.* (1991) noted an improvement in performance in consonant recognition, especially the recognition of voiced stops, when the consonants were amplified by 10 dB. In summary, improving the CVR can potentially improve the intelligibility of some consonants by NH and HI listeners.

In most of the preceding CVR studies, isolated syllables in consonant-vowel (CV) or consonant-vowel-consonant (VCV) format were used as test material, and the amplification was applied to all consonants including the semivowels and nasals. Such an approach raises some questions in terms of practical implications in commercial hearing aids and in terms of generalization of the studies' outcomes and conclusions in continuous speech. For one, it presumes the existence of a consonant detection algorithm that would reliably discriminate between semivowels and vowels or between nasals and vowels, a formidable challenge not only in background noise but also in quiet conditions. Second, in real communicative situations, HI listeners make use of high-level linguistic information (e.g., context) to identify words in continuous

^{a)}Author to whom correspondence should be addressed. Electronic mail: loizou@utdallas.edu

speech. As such, if the information contained in some consonants is masked and not perceptible, listeners might be able to use supplemental cues present in the relatively less corrupted (masked) segments (e.g., vowels). Third, background noise does not typically mask all phonetic segments to the same extent, owing to the spectral differences between noise and speech (Parikh and Loizou, 2005). Low-frequency noise (e.g., car noise), for instance, will not mask the higher frequencies to the same degree as it masks the low-frequency regions of the spectrum. Consequently, a different amplification factor might be required for different consonants. Last, modifying (increasing) the CVR in continuous speech (rather than in isolated syllables) might not always be desirable as it can affect the transmission of voicing information. A study by Port and Dalby (1982), for instance, demonstrated that when other cues to voicing are ambiguous, the CVR can provide reliable cues to perception of voicing of word-final stops. Luce and Charles-Luce (1985) also showed that while vowel duration is very reliable in signaling voicing in syllable-final stops, the CVR remains a significant correlate of voicing. The contribution, and importance, of voicing information to perception of continuous speech in steady-background noise has been demonstrated in our prior study (Li and Loizou, 2008) with NH listeners. Not much work has been done, however, to understand the contribution of voicing information in continuous speech by HI listeners, particularly when speech is corrupted by background noise.

Taking the preceding issues and questions into account, we investigate in the present study a number of hypotheses. The weak consonants, such as the obstruent consonants (e.g., stops, fricatives) are easily masked by noise compared to the more intense vowels and more generally the sonorant sounds (Parikh and Loizou, 2005). This suggests that in a noisy situation, listeners will have access to relatively reliable information contained in the vowels (as they are masked less by noise) and little, if any, information contained in the obstruent consonants. Listeners are thus faced with the task, and challenge, of integrating the information “glimpsed” from the vocalic phonetic segments (owing to their relatively higher SNR) to identify words in continuous speech. The first experiment tests the hypothesis that providing access to information contained in the obstruent consonants ought to improve speech recognition by HI listeners as that would assist them in integrating the information “glimpsed” from the vocalic segments to hear out the target speech. To test this hypothesis, listeners are presented with noise-corrupted sentences containing clean (i.e., not masked by noise) obstruent consonants but otherwise noise-corrupted sonorant sounds (e.g., vowels, semivowels, nasals). This experiment can provide important insights as to how efficiently HI listeners integrate information that is “glimpsed” across multiple segments of the utterance to hear out the target speech. This ability is important particularly in the context of perceiving target speech in competing-talker listening situations (see review by Assmann and Summerfield, 2004; Li and Loizou, 2007). The outcomes of the preceding experiment are important as they would provide motivation for the development of hearing-aid signal processing algorithms that would either provide differential amplification to the heavily masked obstruent consonants

(investigated in Experiment 2) or specific noise reduction techniques capable of suppressing the noise present in those segments. The second experiment investigates the possibility that the high-frequency cues present in most obstruent consonants (e.g., /t/, /s/) might not be audible to HI listeners. To test this hypothesis, noisy consonants contained in sentences were selectively amplified in the high frequencies while leaving the noisy sonorant sounds unaltered. If the high-frequency cues contained in most consonants were not accessible to HI listeners because they were not audible, then amplifying them ought to improve speech intelligibility. The high-frequency consonant amplification was done in a way that did not significantly alter the CVR as the intent was to provide more reliable voicing information.

II. EXPERIMENT 1: PERCEPTION OF CONSONANT INFORMATION BY HI LISTENERS IN CONTINUOUS SPEECH EMBEDDED IN BACKGROUND NOISE

A. Methods

1. Subjects and stimuli

A total of eight HI subjects and eight elderly normal-hearing subjects (ENH) participated in this experiment. HI subjects were 37–78 yr, with average age of 56 yr and the ENH subjects were 48–67 yr with average age of 54 yr. All subjects were native English speakers and were paid for their participation. Figure 1 shows the audiometric thresholds of the HI subjects. All the HI subjects were experienced hearing-aid users and had bilateral sensorineural hearing loss with one exception. One subject had unilateral hearing loss. The inclusion criteria for the ENH subjects was to have maximum hearing thresholds for air conducted pure tones not exceeding 30 dB HL in either of the two ears. Hence, all ENH subjects had thresholds lower than 30 dB HL at octave frequencies from 250 Hz to 4 kHz. Thresholds at 8 kHz were also lower than 30 dB HL with the exception of one subject.

The speech material consisted of sentences taken from the IEEE database (IEEE, 1969). All sentences were

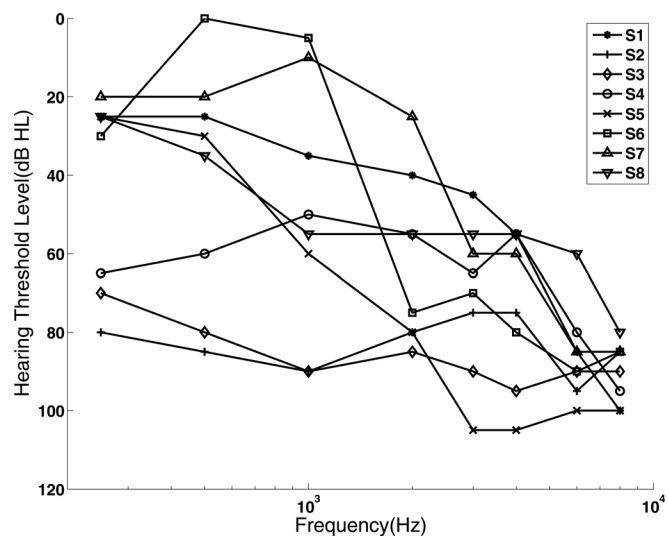


FIG. 1. Audiometric thresholds (in dB HL) of the HI subjects. For the subjects with bilateral hearing loss, only the poorer threshold is shown.

produced by a male speaker. The sentences were recorded in a sound-attenuated booth (Acoustic Systems, Inc.) at a 25 kHz sampling rate. Details about the recording setup and copies of the recordings are available in [Loizou \(2007\)](#). The sentences were corrupted by a 20-talker babble (Auditec CD, St. Louis) at four SNR levels (-5 to $+10$ dB, in steps of 5 dB). The long-term average spectrum of the babble can be found in [Parikh and Loizou \(2005\)](#).

2. Signal processing

The IEEE sentences were manually segmented into two broad phonetic classes: (1) obstruent sounds, which included the stops, fricatives, and affricates, and (2) the sonorant sounds, which included the vowels, semivowels, and nasals. The segmentation was carried out in a two-step process. In the first step, initial classification of voiced and unvoiced speech segments was provided by a highly accurate F0 detector, taken from the STRAIGHT algorithm ([Kawahara et al., 1999](#)), which was applied to the stimuli at 1-ms intervals using a high-resolution fast Fourier transform (FFT). Segments with nonzero F0 values were initially classified as voiced and segments with zero F0 value as unvoiced. In the second step, the voiced and unvoiced decisions were inspected for errors and the detected errors were manually corrected. This process is described in more detail in [Li and Loizou \(2008\)](#). The two-class segmentation of all the IEEE sentences was saved in text files in the same format as TIMIT's .phn files and is available from the CD-ROM provided in [Loizou \(2007\)](#). The preceding two phonetic classes were chosen for two reasons. First, the obstruent sounds are particularly vulnerable to steady background noise, given their low intensity relative to that of the sonorant sounds ([Parikh and Loizou, 2005](#)). Second, the majority of the obstruent sounds have a prominent energy in the high frequencies. The fricatives /s/ and /sh/, for instance, have most of their energy concentrated above 4 kHz.

The IEEE speech stimuli were processed in two different conditions. In the first condition, which served as the control condition, the listeners were presented with noise-corrupted speech stimuli. We refer to this condition as the unprocessed (UN) condition. The second condition included sentences containing clean (uncorrupted) obstruent segments but noise-masked sonorant segments (e.g., vowels). The clean obstruent segments were extracted from the sentences prior to their mixing with the masker. We refer to this condition as the clean consonant (CC) condition.

3. Procedure

Sentences from the IEEE database were processed as described in Sec. II A 2 and stored for testing purposes. Two IEEE lists (20 sentences) were used for each condition, and none of the lists were repeated. Multi-talker babble was added to the IEEE sentences at four different SNR levels: -5 , 0, 5, and 5 dB. HI subjects were tested at all SNR levels, but ENH subjects were only tested at -5 and 0 dB SNR levels as we were constrained by ceiling effects at higher SNR levels. Prior to the test, subjects listened to four lists of sentences to become familiar with the processed stimuli and the

task (i.e., subjects listened to the sentences while reading the contents of the sentence, and this was done only during the practice session). HI subjects participated in a total of eight randomized conditions ($= 4$ SNR levels $\times 2$ algorithms), and ENH subjects participated in a total of four randomized conditions ($= 2$ SNR levels $\times 2$ algorithms). Two lists of IEEE sentences (i.e., 20 sentences) were used per condition, and none of the lists were repeated across conditions. Different lists of sentences were assigned to different listeners. Sentences were presented to the listeners in blocks with 20 sentences/block for each condition. The sentences were presented at an average level of 72 dB SPL as measured by the Quest Precision Impulse Integrating Sound Level Meter (Model No: 1800). During the test, the subjects were asked to type the words they heard on a keyboard.

Because the personal hearing aids might have noise reduction algorithms programmed, the subjects were fit with Phonak eXtra behind-the-ear (BTE) hearing aids that were calibrated according to the subjects' hearing loss and had no noise-reduction capabilities. These BTE hearing aids matched the gain and output of the subjects' personal hearing aids within 5 dB but had no noise-reduction capabilities. Two subjects were fit with eXtra 211 hearing aids, three with eXtra 311 hearing aids, and three with eXtra 411 hearing aids. The hearing aids were programmed based on the NAL-NL1 algorithm using the manufacturer's fitting software ([Byrne et al., 2001](#)). This fitting algorithm is designed to maximize speech intelligibility while keeping the overall loudness at a level that does not exceed that of normal-hearing persons listening to the same sound. The gain required to achieve this varies with input level making it most appropriate for non-linear hearing aids. The subject's own ear-molds were used or a temporary coupling was made using tubing and a foam tip. Real-ear verification according to NAL-NL1 targets was performed using the Audioscan Verifit program and standard probe insertion depth of 5-6 mm from the tympanic membrane. The targets for soft, standard, and loud inputs (50, 70, and 85 dB SPL, respectively) were met within $+5$ dB at octave frequencies from 500 to 4000 Hz. Furthermore, electro-acoustical analysis, according to [ANSI S3.22 \(2003\)](#), showed agreement between high-frequency average maximum output and reference test gain for the experimental and personal aids within $+5$ dB when run at the users' settings. Attack and release times were 1 and 10 ms, respectively. The maximum output and compression settings remained at the default settings as determined by the fitting software. The hearing aids were set to experience level 4, which represents a listener who is accustomed to amplification.

The testing was held in a double-walled IAC sound booth (Model No: 102400). The processed speech files were sent first to a Crown amplifier (Model No: D75) that was used to adjust the presentation level of the signals being played. The Crown amplifier was connected to the speakers in the sound booth. The speech files were played to the subjects using TOA Electric Company speakers (Model No: 22-ME-AV), which have a frequency response up to 20 kHz. For the single HI subject with unilateral hearing loss, the ear with normal hearing was occluded during the test. A computer monitor and a keyboard were placed in the sound booth

so that the subjects could operate a graphical user interface to type in the words they heard using the keyboard. The entire testing was monitored outside the sound booth with an additional computer monitor and keyboard.

B. Results and discussion

Figure 2 shows the mean performance obtained by HI and ENH listeners in the various conditions. Performance was measured in terms of percent of keywords identified correctly. A substantial improvement in intelligibility was observed by ENH listeners at -5 dB SNR when they had access to clean consonant information. Large improvements in intelligibility were also observed by HI listeners at low SNR levels (≤ 0 dB).

Two-way analysis of variance (ANOVA) with repeated measures was run to assess the effects of consonant processing (UN and CC) and SNR level on sentence recognition by the HI listeners. Prior to the ANOVA, the percent correct scores were arcsine transformed, as per Studebaker (1985), to RAU scores. The ANOVA, when applied to the RAU scores, showed significant effect of processing ($F[1,7] = 66.1$, $p < 0.0005$), significant effect of SNR level ($F[3,21] = 197.1$, $p < 0.0005$) and non-significant interaction ($F[3,21] = 2.83$, $p = 0.063$).

The preceding analysis suggests that HI listeners received significant benefit in intelligibility when they had access to clean consonant information in otherwise noise-masked sentences (sonorant segments were left corrupted by babble). The improvement ranged from 20% to 30% for most SNR levels (-5 , 0 , and 5 dB) and was found to be statistically significant ($p < 0.005$) based on *post hoc* tests. Based on interpolation of the psychometric functions (Fig. 2) of the UN and CC conditions, this improvement amounted to a 4-dB decrease in speech reception threshold (SRT). A comparatively lower benefit was obtained at the high-SNR level (10 dB). *Post hoc* tests confirmed that the benefit observed at 10 dB SNR was significant ($p = 0.019$), albeit small (5% improvement). Bonferonni correction was applied to all *post hoc* tests.

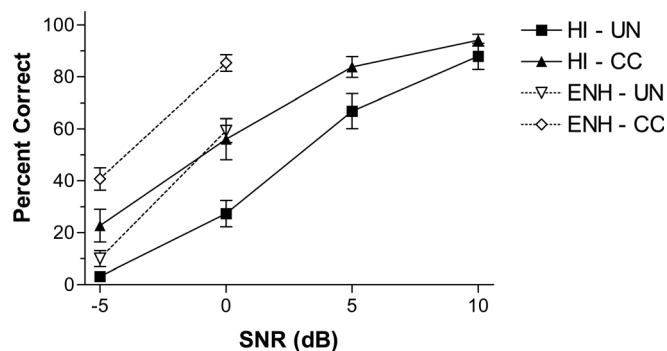


FIG. 2. Mean percent correct scores of HI listeners (dark-filled symbols) and ENH listeners (open symbols) in the various SNR conditions. Performance obtained using unprocessed (noise-masked) sentences is indicated as UN, and performance obtained using sentences containing clean consonants but otherwise noise-masked sonorant sounds is indicated as CC. Error bars indicate standard errors of the mean.

The same statistical analysis was used to analyze the scores obtained by ENH listeners. The ANOVA showed significant effect of processing ($F[1,7] = 110.1$, $p < 0.0005$), significant effect of SNR level ($F[1,7] = 364.3$, $p < 0.0005$) and nonsignificant interaction ($F[1,7] = 1.3$, $p = 0.288$). The preceding analysis suggests that the improvement in intelligibility brought by CC processing was significant at both SNR levels.

It is clear from Fig. 2 that HI listeners benefited when they had access to clean consonant information in otherwise noise-masked sentences. It is worth noting that the improvement in intelligibility with the utilization of CC processing was comparable to that obtained by cochlear implant (CI) listeners (Li and Loizou, 2010). In our previous study, CI listeners were presented with similar speech stimuli (CC) in steady-continuous noise at 5 and 10 dB SNR levels. Results indicated a 20% improvement in intelligibility at 5 dB and a 10% improvement at 10 dB (Li and Loizou, 2010). These improvements are nearly identical to those observed in the present experiment (see Fig. 2) with HI subjects. Hence, from this we conclude that CI listeners are able to extract, and integrate, the consonant information contained in otherwise noise-masked sentences at the same level as HI listeners.

To compare the performance obtained by ENH and HI listeners in the various conditions, a three-way ANOVA was run using the two listener groups, two SNR levels (-5 and 0 dB) and two processing conditions (UN and CC) as factors and the number of words correct as dependent variables. The processing ($F[1,7] = 167.5$, $p < 0.001$) and group effects ($F[1,7] = 15.7$, $p < 0.001$) were found to be significant, and their interaction was found to be non-significant ($F[1,7] = 0.35$, $p = 0.568$). SNR also was found to have a significant effect ($F[1,7] = 216.5$, $p < 0.005$) and the three-way interaction of SNR \times processing \times group was non-significant ($F[1,7] = 5.0$, $p = 0.06$). From the preceding analysis, we can conclude that changes in SNR, processing and group affected performance. Planned comparisons, based on independent samples *t*-tests (2-tailed, unequal variances assumed), between the scores obtained by the two listener groups indicated no significant difference ($p = 0.155$) in benefit from CC processing at -5 dB SNR. There was also no significant difference ($p = 0.686$) in the 0-dB SNR condition.

On average, HI listeners were not able to utilize and integrate the consonant information at the same level as the ENH listeners in the -5 dB SNR condition (HI listeners received only 64% of the benefit obtained by ENH listeners). There was, however, a considerable variability among subjects, and Fig. 3 shows the individual HI listener's scores in the various SNR conditions. As shown in Fig. 3 (SNR = -5 dB), some subjects (S2 and S3) received little benefit, while others received a large benefit (subjects S6 to S8). Subject S7, in particular, received a 60% benefit with CC processing. This variability in performance among HI listeners led us to wonder whether their benefit was somewhat related to their audiometric thresholds. To examine that, we computed the correlation between the CC benefit and average audiometric thresholds (computed using the average thresholds at octave frequencies of 400 Hz to 8000 Hz). The CC benefit was computed as the difference between the score obtained in the CC condition relative to the score obtained in the control UN

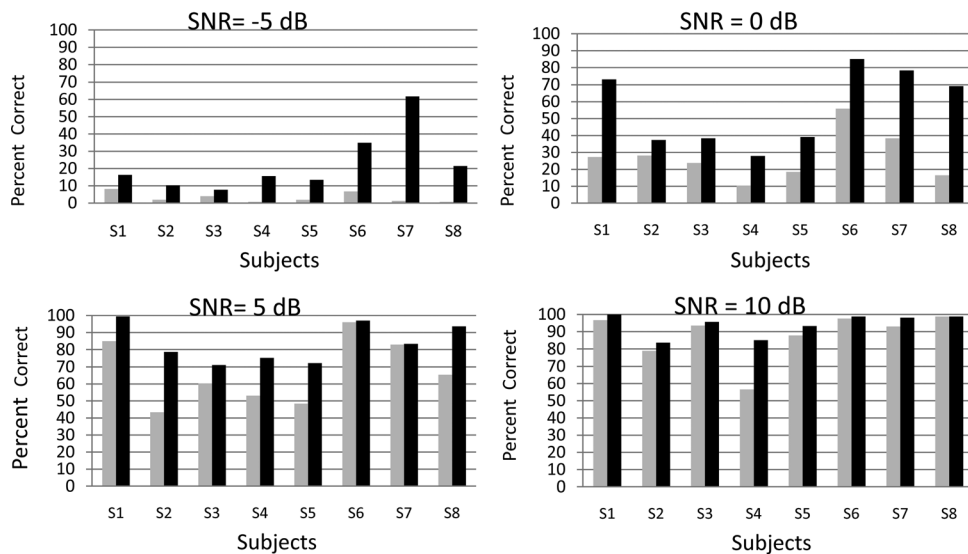


FIG. 3. Individual HI listeners' scores obtained in the unprocessed (noise-corrupted) condition (gray bars) and in the CC condition (dark bars).

condition (no processing). The resulting correlation coefficient was $r = -0.37$ ($p = 0.36$) for SNR = -5 dB and $r = -0.65$ ($p = 0.08$) for SNR = 0 dB. Overall, no significant correlation was observed between audiometric thresholds and benefit with CC processing.

The lack of correlation did not explain why some HI listeners did not receive as much benefit as ENH listeners, at least in the -5 dB SNR condition. There are several possible reasons for that. First, there exists the audibility issue with HI listeners with high-frequency sensorineural hearing loss. As many studies have demonstrated, high-frequency amplification might not be always beneficial, particularly for patients with “dead regions” in their cochlea (Moore, 2001) or for patients with hearing loss exceeding 55 dB HL at high-frequency (≥ 3000 Hz) regions (e.g., Hogan and Turner, 1998; Ching *et al.*, 1998; Turner and Cummings, 1999). We could have increased the presentation level; however, several studies have observed “rollover” effects, in that high presentation levels led to a decrease in speech intelligibility (e.g., Studebaker *et al.*, 1999; Molis and Summers, 2003). In fact, Molis and Summers (2003) have shown that increasing the intensity of the high-frequency speech bands produced more rollover than for the low-frequency bands. To further examine the audibility issue, we analyzed selected high-frequency consonants using short-term acoustic measures, such as those used by Stelmachowicz *et al.* (1993). In particular, 1/3-octave band levels were computed for the consonants /s/ and /t/ taken directly from an IEEE sentence. The 1/3-octave band levels were computed based on the spectrum of 25-ms (Hanning-windowed) segment of the selected consonants, and these are shown in Fig. 4. From these we can deduce that for some HI listeners (e.g., S2, S3) who have low hearing thresholds in the low frequencies (< 1 kHz), portions of the consonant spectra might not have been audible. Furthermore, the frequency location of the spectral peaks of these two consonants was in the range of 5–6 kHz, which might fall outside the bandwidth of most amplification devices.

Second, NH listeners are generally better than most HI listeners at integrating information carried by consonants with otherwise corrupted vowels to understand words in con-

tinuous speech (Kewley-Port *et al.*, 2007). When presented with sentences containing clean consonants and vowels replaced with speech-shaped noise, elderly HI listeners performed worse than young NH listeners (Kewley-Port *et al.*, 2007). On this regard, the outcome of the present study is consistent with that of Kewley-Port *et al.*, (2007). The audibility issue, and to some extent the consonant information integration issue, is examined in the follow up experiment with the use of selective consonant amplification in continuous speech.

III. EXPERIMENT 2: SELECTIVE AMPLIFICATION OF CONSONANTS IN CONTINUOUS SPEECH CORRUPTED BY BACKGROUND NOISE

The data from Experiment 1 indicated that some HI subjects were not able to perceive (and integrate) the information carried by consonants in continuous speech to the same extent

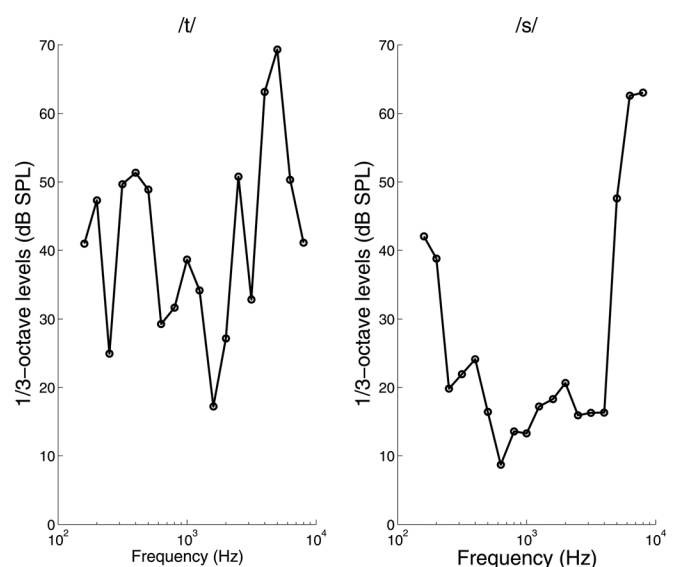


FIG. 4. 1/3-octave band levels (dB SPL) as a function of frequency for segments of the consonants /t/ (left panel) and /s/ (right panel) taken from an IEEE sentence.

as ENH listeners did, at least in the extremely low SNR (-5 dB) condition. The reasons for that were unclear. In this experiment, we investigate the hypothesis that the high-frequency cues present in consonants (e.g., /t/, /s/) were not audible to them and thus not perceptible. To test this hypothesis, we selectively amplified the noisy consonants contained in sentences while leaving the noisy sonorant sounds unaltered. If the consonants were somehow not audible to HI listeners at low SNR levels, then amplifying them ought to reduce the performance difference between normal-hearing and HI listeners. This hypothesis is tested in the present experiment. Amplifying the consonants (while leaving the level of the sonorant sounds the same) will alter the CVR, and we will thus be in a position to assess the impact of CVR modification on continuous speech recognition.

A. Methods

1. Subjects and stimuli

The same subjects who participated in Experiment 1 also participated in the present experiment. Sentences taken from the IEEE corpus (available in [Loizou, 2007](#)) were used as test

material. The sentence lists chosen for the present experiment differed from those used in Experiment 1. The same masker (20-talker babble) was used as in Experiment 1.

2. Signal processing

The IEEE sentences were manually segmented, as before, into obstruent and sonorant sounds. Segments that were classified as sonorants were left unaltered, i.e., remained corrupted by the background noise (20-talker babble). Segments that were classified as obstruent sounds were processed as follows. The fast Fourier transform (FFT) was computed for each 10-ms frame (with 2-ms update) of the obstruent sounds, and spectral components falling at frequencies of 1.5 kHz or above were amplified by a factor of 2 (6.02 dB amplification), while spectral components at frequencies below 1.5 kHz were attenuated by a factor of 0.1 (20 dB attenuation). The attenuation of the low-frequency region (0–1.5 kHz) was introduced for the following two reasons: (1) to make the vowel/consonant boundaries more distinct because the steady background noise tends to blur the low-frequency acoustic landmarks and (2) many of the obstruent sounds (e.g., /s/, /t/) inherently have low energy in the low frequencies. Following the selective

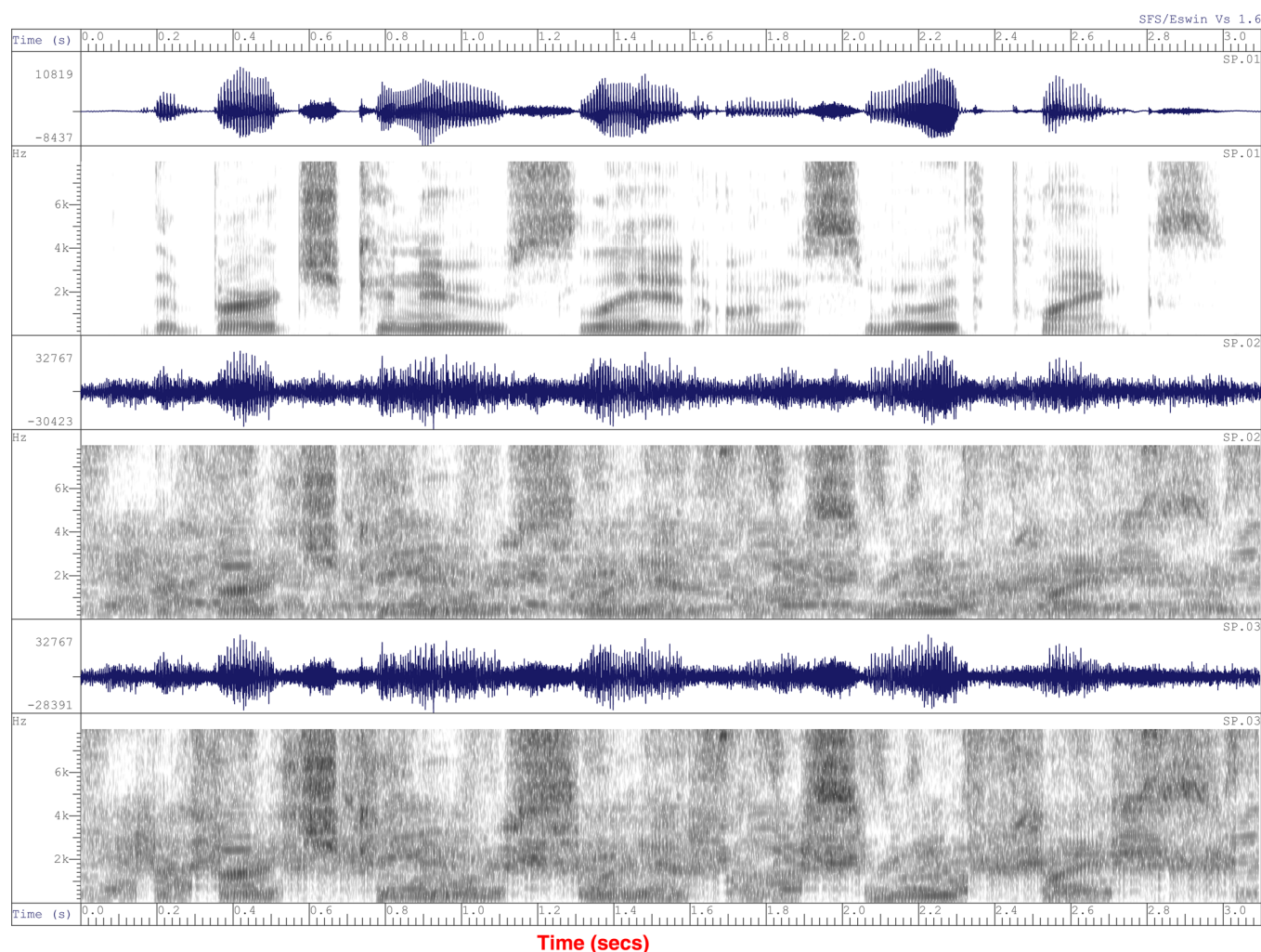


FIG. 5. (Color online) Waveforms and wide-band spectrograms of an IEEE sentence in quiet (top panel), sentence corrupted at 0 dB SNR (middle panel), and AC-processed (bottom panel) such that the consonants were amplified in the high-frequency regions (> 1.5 kHz) and attenuated in the low-frequency region (< 1.5 kHz).

consonant amplification/attenuation, which was implemented in the frequency domain, an inverse FFT was taken to reconstruct the signal in the time domain. The phase spectrum of the noisy speech was used during synthesis. The synthesized signal was finally normalized so that the overall RMS level of the processed signal was the same as that of the noisy signal. We henceforth denote the preceding processing as AC (amplification of consonants) processing.

Figure 5 shows example spectrograms of AC processing. In this example, the sentence was corrupted by babble at 0 dB SNR. Two distinct effects of AC processing are evident from this figure: (1) the consonants with high-frequency prominence are accentuated, and perhaps become more audible to HI listeners, and (2) the voiced/unvoiced boundaries distinguishing vowels from consonants are also more evident compared to those present in the unprocessed sentence. The latter effect is produced by the low-frequency attenuation of consonants. It is also worth noting that the CVR decreased, rather than increased, and was caused by the low-frequency attenuation of consonants.

The amplification and attenuation factors (2 and 0.1, respectively) used were chosen based on a pilot study conducted with a few ENH and HI listeners. The amplification factor was varied in the range of 1.1 to 10 and applied to sets of 20 sentences. Listening tests indicated that values below 2 did not make any notable difference in intelligibility, whereas values greater than 4 produced distortions in the synthesized signal. Similarly, the attenuation factor was varied in the range of 0.01 to 0.3 and applied to sets of 20 sentences. The value of 0.1 was found to yield the highest intelligibility benefit (relative to the intelligibility of unprocessed stimuli) and was thus adopted in this experiment.

When selective amplification was applied to the weak consonants (e.g., /t/), while leaving the intense sonorant sounds (e.g. vowels, semivowels) intact, the resulting consonant-to-vowel energy ratio (CVR) changed. To assess whether the CVR of the modified speech stimuli improved (reduced) after the selective consonant amplification, we computed the CVR at different input SNR levels and compared the CVR levels of the AC processed and unprocessed speech stimuli. The overall CVR was computed by dividing the sum of the power of all unvoiced segments within the utterance by the power of all voiced segments. Unlike the CVR measurements made by Freyman *et al.* (1991), we included the stop closures in the computation of the cumulative power of all unvoiced segments. The resulting CVR values are provided in Table I. These values were obtained based on the average CVR values computed across 10 IIEEE sentences. From this table, we observe that the CVR did improve (i.e., decreased in value) by about 3 dB at low SNR levels and by 1–2 dB at higher SNR levels. We consider the decrease in CVR an improvement as it more accurately reflects the natural intensity ratios occurring in sentences in quiet. In contrast, in noisy conditions, noise masked the silent (closures) segments of the unvoiced segments, i.e., it filled up the waveform valleys, and thus reduced the inherent difference in intensity (energy) between the vowels and consonants. A CVR value close to 0 dB, for instance, would suggest that the vowels and consonants had roughly equal

TABLE I. Measurements (in dB) of consonant-to-vowel ratios (CVRs) for sentences in quiet, noise-corrupted sentences at –5 to 10 dB SNR, and AC-processed sentences. Mean and standard deviations (STDs) of CVRs obtained using 10 IIEEE sentences are reported. The last column shows the mean decrease (dB) in CVR following AC processing relative to the corresponding CVR of noise-corrupted sentences.

SNR	CVRs (dB)	Clean sentences	Noisy unprocessed	AC processed	Decrease re: noisy CVR
–5	Mean	–19.44	–2.41	–5.97	3.56
	STD	2.02	0.84	0.92	
0	Mean	–17.39	–4.64	–7.62	2.98
	STD	1.95	0.91	1.22	
5	Mean	–15.58	–7.36	–9.42	2.07
	STD	3.40	1.55	2.51	
10	Mean	–16.25	–11.20	–12.20	1.00
	STD	2.05	1.35	2.06	

intensity. Such values are indeed observed at extremely low SNR levels (e.g., SNR = –5 dB in Table I) for the noisy speech stimuli.

3. Procedure

The same procedure taken in Experiment 1 was followed in the present experiment. The testing was carried out in two to three sessions (lasting a total of 4–5 h) for the HI listeners and in a single session for the ENH subjects (1–2 h).

B. Results and discussion

Figure 6 shows the mean intelligibility scores obtained by HI and ENH listeners when selective amplification (AC processing) was applied to the noisy sentences. Figure 7 shows the individual HI listeners' scores obtained in the various conditions. The UN scores were taken from Experiment 1. As shown in Fig. 7, AC processing produced small improvements in intelligibility for HI listeners at the low SNR conditions (–5 and 0 dB). No improvement was noted at higher SNR levels. No improvement in intelligibility was noted for the ENH listeners at any SNR condition.

Two-way ANOVA, with repeated measures, was run to assess the effects of consonant processing (UN and AC) and SNR level on sentence recognition by HI listeners. The ANOVA, when applied to the RAU scores (Studebaker, 1986), showed a significant effect of processing ($F[1,7] = 9.7$, $p = 0.017$), significant effect of SNR level ($F[3,21] = 162.9$, $p < 0.0005$) and significant interaction ($F[3,21] = 4.8$, $p = 0.01$). The interaction was caused by the fact that the improvement obtained with selective consonant amplification (AC) was dependent on the SNR level. *Post hoc* tests indicated small, but statistically significant ($p < 0.05$) improvements in intelligibility with AC processing at low SNR levels (–5 and 0 dB). No significant improvement ($p > 0.2$) was observed, however, when AC processing was applied at higher SNR levels (5 and 10 dB).

A similar two-way ANOVA was also run on the scores obtained by ENH listeners. The ANOVA showed a nonsignificant effect of processing ($F[1,7] = 5.47$, $p = 0.052$), significant effect of SNR level ($F[1,7] = 200.4$, $p < 0.0005$) and

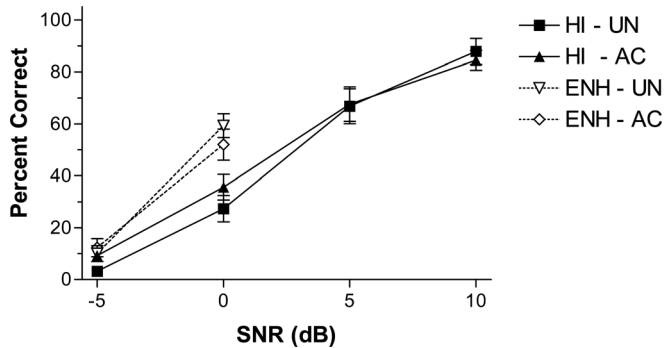


FIG. 6. Mean percent correct scores of HI listeners (dark-filled symbols) and ENH listeners (open symbols) in the various SNR conditions. Performance obtained using unprocessed (noise-corrupted) sentences is indicated as UN, and performance obtained using sentences wherein the consonants were amplified is indicated as AC. Error bars indicate standard errors of the mean.

nonsignificant interaction ($F[1,7] = 3.4, p = 0.106$). This analysis suggests that ENH listeners received no benefit from AC processing.

Small, yet statistically significant, improvement ($\sim 10\%$) was obtained by HI listeners when selective consonant amplification was applied to the noisy sentences. There are several factors that might have contributed to this small improvement. First, following AC processing, the CVR improved (decreased) by as much as 3 dB (Table I). This change in CVR, however, was small and likely not perceptible. In the study by Freyman *et al.* (1991), for instance, a 10-dB change in CVR was found necessary to observe any improvement in intelligibility of isolated VCV stimuli. Second, high-frequency amplification of consonants might have made the consonants slightly more audible to the HI listeners. While this might be the case for some of the consonants, the high-frequency amplification might have also provided conflicting or incorrect cues to consonant identification. This is so because the high-frequency amplification, as applied in the present experiment, is done consistently for all noise-corrupted (obstruent) consonants without paying attention as

to whether the consonant had originally a high-frequency energy prominence (e.g., /s/, /t/, /z/), a low-frequency energy prominence (e.g., /b/, /p/), or a mid-frequency prominence (e.g., /g/) or had a flat spectrum (e.g., /f/, /v/). Taking a look at the example in Fig. 5, we see that consonant amplification is not necessary (or desirable) at segments near $t = 2.4$ s containing stop closures, but it is necessary at segments near $t = 1.9$ s and $t = 2.2$ s, which have naturally high-frequency energy. Finally, the low-frequency attenuation of consonants provided (at least indirectly) access to somewhat reliable acoustic landmarks signifying the onsets/offsets of words or syllables. Such landmarks might facilitate better lexical access and subsequently improve word recognition (Stevens, 2002; Li and Loizou, 2008). In summary, we believe that the small improvement in intelligibility obtained by AC processing can be attributed to the fact that the weak consonants became more audible—at least those with naturally high-frequency prominence in energy, e.g., /s/, /t/, /d/, /z/, /sh/. In addition, we cannot exclude the possibility that listeners had access to relatively reliable low-frequency acoustic landmarks and subsequently better voicing information. This speculation was based on the findings from our prior study with normal-hearing listeners (Li and Loizou, 2008). Access to low-frequency acoustic landmark information signaling the onset of stops or fricatives improved the transmission of voicing information.

In terms of practical implications, the present study showed that small improvements in intelligibility are indeed possible with the use of selective consonant amplification provided that the voiced/unvoiced distinction can be made automatically, and reliably, based solely on information available from the noisy speech stimuli. That is, a simple signal processing algorithm can be devised that first identifies the voiced/unvoiced boundaries and then amplifies/attenuates accordingly the high/low frequency regions of the consonant spectra. A number of voiced/unvoiced detection algorithms can be used, for instance, to detect the voiced/unvoiced boundaries (see review in Konzo, 2004). Some of these algorithms

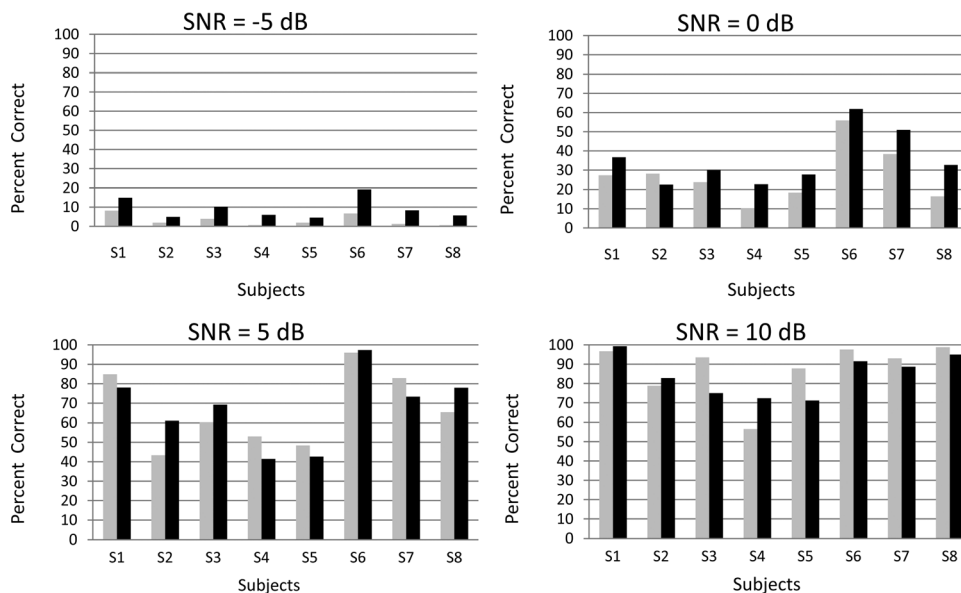


FIG. 7. Individual HI listeners' scores obtained in the unprocessed (noise-corrupted) condition (gray bars) and in the AC condition (dark bars).

exploit well-known characteristics of voiced and unvoiced sounds to make the distinction. These distinctive characteristics include speech periodicity (or lack thereof), peakiness of speech, number of zero crossings of the amplitude waveform (voiced segments have typically a smaller number), and low-frequency to full-spectrum energy ratio among others. A higher benefit in intelligibility is anticipated if noise reduction algorithms were to be applied to the detected consonant segments to suppress the background noise present. As demonstrated in Experiment 1, improvements as large as 30% in intelligibility (4 dB improvement in SRT) are to be expected on the average. Some subjects might receive considerably larger benefit (e.g., subjects S7 and S8 in Fig. 3).

In the context of designing noise-reduction algorithms for hearing aids, the present study showed the potential of selective consonant amplification at low SNR levels. The surprising finding of our study is that small improvements (~10%) in intelligibility were observed in the absence of applying any noise-suppression algorithm on the identified (noise-corrupted) consonant segments. Further improvements in intelligibility are thus to be expected if specific noise reduction algorithms are applied to somehow “enhance” the consonant segments alone. It should be noted that existing modulated-based noise reduction algorithms used in commercial hearing aids fail to produce any significant improvements in intelligibility (Bentler *et al.*, 2008). And most noise-reduction algorithms used in hearing aids (Bentler and Chiou, 2006) are applied to both voiced (e.g., vowels) and unvoiced segments (e.g., stops) of the utterance. The outcomes from Experiments 1 and 2 taken together, however, suggest that applying noise suppression only to the (obstruent) consonant segments of the utterance ought to be sufficient in terms of obtaining large improvements (4-dB decrease in SRT) in intelligibility. Further work is warranted to assess the effectiveness of noise-reduction algorithms in suppressing the masking (by noise) of the weak consonants.

IV. SUMMARY AND CONCLUSIONS

The present study assessed the contribution of information carried by obstruent consonants to speech intelligibility in noise by HI listeners. Based on the outcomes of the two experiments, the following conclusions can be drawn:

- (1) HI listeners received significant benefits in intelligibility when they had access to clean consonant (obstruent sounds) information in otherwise noise-masked sentences (sonorant segments were left corrupted by background noise). The improvement ranged from 20% to 30% for most SNR levels (−5, 0, and 5 dB), and amounted to a 4-dB decrease in speech reception threshold (Fig. 2).
- (2) At extremely low SNR levels (−5 dB), HI listeners received about 64% of the benefit obtained by ENH listeners (Fig. 2). There was, however, considerable variability among subjects (Fig. 3).
- (3) Listening tests (Fig. 6) indicated that small (~10%), but statistically significant, improvements in intelligibility were obtained by HI listeners at low SNR conditions (−5 and 0 dB) when the consonants were amplified by

6 dB in the high-frequency regions (>1.5 kHz) and attenuated by 20 dB in the low-frequency regions. This form of selective consonant amplification produced a small decrease (2-3 dB) in the consonant-to-vowel ratio (see Table I) and provided reliable low-frequency acoustic landmarks (e.g., vowel/consonant boundaries) signifying the onsets of syllables or words (see example in Fig. 5). This in turn facilitated a better lexical segmentation of the speech stream (Stevens, 2002) leading to the small improvement in intelligibility.

In the context of designing noise-reduction algorithms for hearing aids, the present study demonstrated the potential of selective consonant amplification. That is, a simple signal processing algorithm can be devised that first identifies the voiced/unvoiced boundaries and then amplifies/attenuates accordingly the high/low frequency regions of the consonant spectra.

ACKNOWLEDGMENTS

This work was completed by the first author (R.S.) as part of her Master’s thesis requirement in the Department of Electrical Engineering. The work was supported by Grant No. R01 DC010494 (P.C.L) awarded from the National Institute of Deafness and other Communication Disorders (NIDCD) of the National Institutes of Health (NIH). The authors would like to thank Dr. Fei Chen for all his help with one of the experiments.

ANSI (2003). S3.22, *Specification of Hearing Aid Characteristics* (Acoustical Society of America, New York).

Assmann, P. F., and Summerfield, Q. (2004). “The perception of speech under adverse acoustic conditions,” in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York), pp. 231–308.

Bentler, R. A., and Chiou, L-K. (2006). “Digital noise reduction: An overview,” *Trends Amplif.* **10**, 67–83.

Bentler, R., Wu, Y., Kettel, J., and Hurtig, R. (2008). “Digital noise reduction: Outcomes from laboratory and field studies,” *Intern. J. Audiol.* **47**, 447–460.

Byrne, D., Dillon, H., Ching, T., Katsch, R., and Keidser G. (2001). “NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures,” *J. Am. Acad. Audiol.* **12**(1), 37–51.

Ching, T. Y., Dillon, H., and Byrne, D. (1998). “Speech recognition hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification,” *J. Acoust. Soc. Am.* **103**, 1128–1140.

Edwards, B. (2004). “Hearing aids and hearing impairment,” in *Speech Processing in the Auditory System*, edited by S. Greenberg (Springer Verlag, New York), pp. 339–421.

Freyman, R. L., Nerbonne, G. P., and Cote, H.A. (1991). “Effect of consonant-vowel ratio modification on amplitude envelope cues for consonant recognition,” *J. Speech Hear. Res.* **34**, 415–426.

Freyman, R. L., and Nerbonne, G. P. (1989). “The importance of consonant-vowel intensity ratio in the intelligibility of voiceless consonants,” *J. Speech Hear. Res.* **32**, 524–535.

Gordon-Salant, S. (1986). “Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing,” *J. Acoust. Soc. Am.* **80**, 1599–1607.

Gordon-Salant, S. (1987). “Effects of acoustic modification on consonant recognition by elderly hearing-impaired listeners,” *J. Acoust. Soc. Am.* **81**, 1199–1202.

Hogan, C., and Turner, C. (1998). “High-frequency audibility: Benefits for hearing-impaired listeners,” *J. Acoust. Soc. Am.* **104**, 432–441.

IEEE. (1969). “IEEE recommended practice speech quality measurements,” *IEEE Trans. Audio Electroacoust.* **AU-17**, 225–246.

Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A. (1999). “Restructuring speech representations using a pitch-adaptive time-

- frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," *Speech Comm.* **27**, 187–207.
- Kewley-Port, D., Burkle, Z., and Lee, J. (2007). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," *J. Acoust. Soc. Am.* **122**, 2365–2375.
- Kondo, A. (2004). *Digital Speech*, 2nd ed. (Wiley, West Sussex, UK), Chap. 6, pp. 149–197.
- Li, N., and Loizou, P. (2007). "Factors influencing glimpsing of speech in noise," *J. Acoust. Soc. Am.* **122**, 1165–1172.
- Li, N., and Loizou, P. (2008). "The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise," *J. Acoust. Soc. Am.* **124**(6), 3947–3958.
- Li, N., and Loizou, P. (2010). "Masking release and the contribution of obstruent consonants on speech recognition in noise by cochlear implant users," *J. Acoust. Soc. Am.* **128**(3), 1262–1271.
- Loizou, P. C. (2007). *Speech Enhancement: Theory and Practice* (CRC Press, Boca Raton, FL), Appendix C, pp. 589–599.
- Luce, P., and Charles-Luce, J. (1985). "Contextual effects on vowel duration, closure duration and the consonant/vowel ratio in speech production," *J. Acoust. Soc. Am.* **78**(6), 1949–1957.
- Montgomery, A. A., and Edge, R. (1988). "Evaluation of two speech enhancement techniques to improve intelligibility for hearing-impaired adults," *J. Speech Hear. Res.* **31**, 386–393.
- Molis, M., and Summers, V. (2003). "Effects of high presentation levels on recognition of low- and high-frequency speech," *ARLO* **4**, 124128.
- Moore, B. C. J. (2001). "Dead regions in the cochlea: Diagnosis, perceptual consequences, and implications for the fitting of hearing aids," *Trends Amplif.* **5**, 1–34.
- Owens, E., Talbot, C., and Schubert, E. (1968). "Vowel discrimination of hearing impaired listeners," *J. Speech Hear. Res.* **11**, 648–655.
- Owens, E., Benedict, M., and Schubert, E. (1972). "Consonant phonemic errors associated with pure-tone configurations and certain kinds of hearing impairment," *J. Speech Hear. Res.* **15**, 308–322.
- Parikh, G., and Loizou, P. (2005). "The influence of noise on vowel and consonant cues," *J. Acoust. Soc. Am.* **118**, 3874–3888.
- Port, R., and Dalby, J. (1982). "Consonant/vowel ratio as a cue of voicing in English," *Percept. Psychophys.* **32**, 141–152.
- Skinner, M. W. (1980). "Speech intelligibility in noise-induced hearing loss: Effects of high-frequency compensation," *J. Acoust. Soc. Am.* **67**(1), 306–317.
- Stelmachowicz, P., Mace, A., Kopun, J., and Carney, E. (1993). "Long-term and short-term characteristics of speech: Implications for hearing-aid selection for young children," *J. Speech Hear. Res.* **36**, 609–620.
- Stevens, K. N. (2002). "Towards a model of lexical access based on acoustic landmarks and distinctive features," *J. Acoust. Soc. Am.* **111**(4), 1872–1891.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Studebaker, G., Sherbecoe, R., McDaniel, M., and Gwaltney, C. (1999). "Monosyllabic word recognition at higher-than-normal speech and noise levels," *J. Acoust. Soc. Am.* **105**, 2431–2444.
- Turner, C. W., and Cummings, K. J. (1999). "Speech audibility for listeners with high-frequency hearing loss," *Am. J. Audiol.* **8**, 47–56.
- Turner, C. W., and Robb, M. P. (1987). "Audibility and recognition of stop consonants in normal and hearing-impaired subjects," *J. Acoust. Soc. Am.* **81**, 1566–1573.
- Yund, E., and Buckles, K. (1995). "Multichannel compression in hearing aids: Effect of number of channels on speech discrimination in noise," *J. Acoust. Soc. Am.* **97**, 1206–1223.