Running head: INTELLIGIBILITY OF MODIFIED SPEECH

The Intelligibility of Modified Speech for Young Listeners with Normal and Impaired Hearing

Rosalie M. Uchanski and Ann E. Geers
Central Institute for the Deaf, Saint Louis, Missouri
Athanassios Protopapas
Scientific Learning Corporation, Berkeley, California

Abstract

Exposure to modified speech has been shown to benefit language-learning impaired children with respect to their language skills (Tallal et al., 1996; Merzenich et al., 1998). In the study by Tallal and colleagues, the speech modification consisted of both slowing down and amplifying fast, transitional elements of speech. In this study, we examined whether the benefits of modified speech could be extended to provide intelligibility improvements for children with severe-to-profound hearing impairment who wear sensory aids. In addition, the separate effects on intelligibility of slowing down and amplifying were evaluated.

Two groups of listeners were employed: eight severe-to-profoundly hearing-impaired children and five children with normal hearing. Four speech-processing conditions were tested: 1) natural, unprocessed speech; 2) envelope-amplified speech; 3) slowed speech; and 4) both slowed and envelope-amplified speech. For each condition, three types of speech materials were used: words in sentences, isolated words, and syllable-contrasts. To degrade the performance of the normal-hearing children, all testing was completed with a noise background.

Results from the hearing-impaired children showed that all varieties of modified speech yielded either equivalent or poorer intelligibility than unprocessed speech. For words in sentences and isolated words, the slowing-down of speech had no effect on intelligibility scores while envelope-amplification, both alone and combined with slowing-down, yielded significantly lower scores. Intelligibility results from normal-hearing children listening in noise were somewhat similar to those from hearing-impaired children. For isolated words, the slowing-down of speech had no effect on intelligibility while envelope-amplification degraded intelligibility. For both subject groups, speech processing had no statistically significant effect on syllable discrimination. In summary, without extensive exposure to the speech processing conditions, children with

impaired hearing and children with normal hearing listening in noise received no intelligibility advantage from either slowed speech or envelope-amplified speech.

The Intelligibility of Modified Speech for Young Listeners with Normal and Impaired Hearing

At oral schools for the deaf, such as Central Institute for the Deaf (CID), there is an obvious, critical need for speech to be delivered more intelligibly to children with impaired hearing. Even with the most advanced hearing aids or cochlear implants, severely and profoundly hearing impaired children often have great difficulty perceiving speech (Fryauf-Bertschy, Tyler, Kelsay, Gantz & Woodworth, 1997). Such children with severe and profound hearing impairment may perceive correctly only 26% of the words presented to them (Kirk, Pisoni & Osberger, 1995). Our primary interest in studying a speech modification, one introduced recently by Tallal et al. (1996) for language-learning impaired (LLI) children, stems from this critical need to make speech more intelligible for severe-to-profoundly hearing-impaired children.

Intensive exposure to modified speech, as introduced by Tallal et al. (1996), is a major component of their multi-week training program designed for LLI children. Merzenich et al. (1998) and Tallal et al. (1996) report that this multi-week training program provides substantial benefit to LLI children with respect to their language scores. In the earlier study, Tallal et al. (1996) employed two groups of LLI children: 1) a test group which received the four-week intensive training program with modified speech; and 2) a control group which received the four-week intensive training program with unmodified speech. During this intensive program, speech occurred in the context of several computer-based training exercises. After completion of the program, both the control and test groups showed improvements in speech and language scores. However, the test group, which was exposed to modified speech, achieved significantly greater gains than the control group, which was exposed to unmodified speech. Thus, it appears that the speech modification of Tallal et al. contributes to the observed language benefit beyond that which is achieved solely from the computer-based exercises.

The type of speech modification employed by Tallal et al. is motivated by the hypothesis that LLI children “have a ‘temporal processing deficit’ expressed by limited abilities at identifying some brief phonetic elements presented in specific speech contexts and by poor performances at identifying or sequencing short-duration acoustic stimuli presented in rapid succession” (Merzenich et al., 1996, p.77). Their speech modification addresses this deficit in two ways. First, speech is uniformly slowed down by as much as 50% and second, fast-varying elements of the speech signal are amplified. The latter modification is referred to here as envelope amplification (Nagarajan et al., 1998). Both components of the modification are designed to enhance temporally-short, time-varying speech elements, such as the formant transitions from a consonant to a vowel.

Though the speech modification employed by Tallal et al. was intended for LLI children, there are good reasons for exploring the effects of this speech modification for a different population, namely severe-to-profoundly hearing-impaired children (with no other handicapping condition) wearing sensory aids. As mentioned previously, children with severe and profound hearing impairments do not perceive speech well even when aided. Hence, any type of speech processing that might conceivably enhance the intelligibility of speech to these listeners should be explored for potential benefit. Additionally, for hearing-impaired children, better speech perception scores are often associated with better spoken language skills (Boothroyd, Geers & Moog, 1991; Geers & Moog, 1992). So, a benefit in speech perception could also have a positive impact on the language skills of hearing-impaired children.

The envelope-amplification component of Tallal’s speech modification, in particular, appears promising for its potential to improve speech perception. Recently, Hazan and Simpson (1998) reported that explicit amplification of consonants and their subsequent formant transitions[1] improved speech intelligibility in noise for listeners with normal hearing. Thus, if the envelope-amplification described by Nagarajan et al. (1998) does indeed amplify formant transitions while not introducing concomitant degradations, we might expect envelope-amplification to improve the intelligibility of speech for hearing-impaired listeners. Also, little is known about the effects of envelope-amplification on speech perception. This specific type of processing has not been studied and is not comparable to other types of processing, such as amplitude compression, that have been examined extensively (e.g., Moore, Peters & Stone, 1999; Plomp, 1988).

Over the years, the effects of time-expanded speech on intelligibility have been explored in young normal-hearing listeners (Schon, 1970; Korabic, Freeman & Church, 1978), in aged normal-hearing listeners (Schon, 1970; Schmitt, 1983; Gordon-Salant, 1986), in hearing-impaired listeners (Picheny, Durlach & Braida, 1989; Uchanski, Choi, Braida, Reed & Durlach, 1996), and in language-impaired or dyslexic listeners (Stollman, Kapteyn & Sleeswijk, 1994; McAnally, Hansen, Cornelissen & Stein, 1997). Despite many differences amongst these studies (such as the language used, speech materials, listener characteristics, and the amount of time-expansion) there is general agreement that time-expansion does not significantly affect speech intelligibility. That is, time-expansion (by 50% and more) neither degrades nor improves speech intelligibility.

The only studies that showed an improvement in intelligibility for a time-expanded speech signal were those that examined naturally produced clear speech. In these studies, an intelligibility advantage was found for naturally produced clear speech relative to conversational speech, for hearing-impaired adults and for normal-hearing listeners in noise (Picheny, Durlach & Braida, 1985; Payton, Uchanski & Braida, 1994; Uchanski et al., 1996). While clear speech is generally produced at a slower speaking rate (approximately 90-100 wpm for clear as compared to 160-200 wpm for conversational speech), there is growing evidence that clear speech is not equivalent to either naturally-produced slow speech or artificially time-expanded conversational speech. All these types of speech (natural clear, natural slow, artificially time-expanded) differ significantly in intelligibility and in many acoustic properties other than duration (Moore & Zue, 1985; Picheny, Durlach & Braida, 1986; Moon & Lindblom, 1994; Krause, 1995; Fosler-Lussier & Morgan, 1999).

The effect of time-expansion on the ability to discriminate speech sounds is somewhat different from its effect on speech intelligibility or identification. For example, for listeners with normal hearing discriminating sounds in a [ba]-[da] continuum, Sussman and Carney (1989) found no effect of transition duration for 7- to 8-year-old children and a significant effect of transition duration for adults, 5- to 6-year-old children, and 9- to 10-year-old children. For children with language disabilities, slowing down formant transitions consistently improves discrimination between synthetic speech sounds (Alexander & Frost, 1982; Tallal & Piercy, 1975) and seems to enhance the neural representations of synthetic /da/'s and /ga/'s (Bradlow et al., 1999).

We hypothesized that modified speech, with its presumably more salient speech sounds for LLI children, might be more intelligible than unmodified (unprocessed) speech for children with impaired hearing who wear hearing aids and/or cochlear implants. To test this hypothesis we examined the intelligibility of modified speech for children with impaired hearing. The speech modification applied by Tallal et al. (1996), known to be beneficial for training with LLI children, included both envelope-amplification and time-expansion, and thus it should preferably be evaluated as such. On the other hand, as discussed above, previous research with hearing-impaired persons indicated that time-expansion alone was unlikely to increase intelligibility, whereas envelope-amplification might be more successful. Because it is not possible to predict the effect of time-expansion in combination with envelope-amplification, we chose to evaluate all possible modification conditions. That is, for this study the two speech modification components, time-expansion and envelope-amplification, are evaluated separately for their effects on speech intelligibility. Additionally, there is a practical reason for determining these separate effects. A real-time implementation of envelope-amplification would preserve the natural synchrony between the visual and auditory signals that is critical for speechreading by hearing-impaired individuals. By contrast, time-expansion would destroy this natural synchrony between the visual and auditory speech signals.

Besides examining the effect of time-expansion and envelope-amplification on *intelligibility*, the effects of these modifications on *speech discrimination* were also examined. We chose to include a speech-discrimination task because of the promising results from studies of time-expansion on synthetic speech discrimination, and because it is possible for a speech modification to improve the perceptual discrimination of speech sounds *without* improving overall intelligibility. Thus, inclusion of this task allows another opportunity for uncovering a potential perceptual benefit from any of the speech modifications.

While the primary goal of this study was to determine the intelligibility benefit of modified speech for children with hearing-impairment, a group of children with normal hearing were also tested. Tests with hearing-impaired children allowed us to assess the potential benefit of modified speech on intelligibility directly for this population of interest. However, tests with normal-hearing children (listening in noise to eliminate ceiling effects in performance) allowed us to assess the general effect of modified speech on intelligibility, for children with normal auditory processing skills. Also, speech presented to normal-hearing listeners will be affected only by the signal processing of the Tallal speech modification whereas speech presented to hearing-impaired listeners will be affected by the signal processing in the speech modification and by the signal processing (such as a compression algorithm) in the listener's prosthetic hearing device. Consequently, speech perception results from hearing-impaired listeners might be confounded by an interaction between the two types of signal processing while speech perception results from normal-hearing listeners will not.

Method

Participants

Two groups of children participated in this study. For the first group, children with bilateral, sensorineural hearing impairment were recruited from the CID school. All children at CID's school who achieved a minimum score of 5 years on the receptive portion of the Peabody Picture Vocabulary Test (Dunn & Dunn, 1981) were recruited. Receptive vocabulary (PPVT) was used as the primary selection criterion to ensure that participants possessed a vocabulary level appropriate for the speech materials employed in the experiments. A total of eight children with impaired hearing agreed to participate in the study. In addition, the non-verbal cognitive function of these children was tested and determined to be in the normal range for their chronological age. Table 1 lists characteristics of these children such as pure-tone-average, type of hearing device(s) used, age, and PPVT score. As shown in Table 1, a variety of losses, devices, ages, and equivalent receptive language-ages (as based on PPVT) are represented in this group. Six children have profound hearing loss (hi1-hi5, hi7), one child has a moderate-to-severe loss (hi6), and one has normal hearing below 500 Hz with a sloping-to-moderate loss at 1000-8000 Hz (hi8). Four children wear cochlear implants, three wear hearing aids, and one child wears both a hearing aid and cochlear implant (hi7). Four of the five cochlear implants were programmed with the SPEAK processing strategy while one employed the MPEAK processing strategy (hi4). The hearing aids worn by these listeners also varied. The hearing aids worn by two subjects (hi7, hi8) used linear amplification with peak-clipping while the aids worn by others employed wide-band amplitude compression (hi5, hi6).

The second group consisted of five children with normal hearing. These participants were recruited from parents on the CID staff. They ranged in age from 7 to 11 years old, spoke English as their native language, and had normal hearing. Though the PPVT was not employed for the children with normal hearing, there were no known language impairments. On average, the normal-hearing group was younger (mean age: 9 years 5 months) than the hearing-impaired group (mean age: 12 years 4 months), but presumably had a higher mean PPVT age. Despite the difference in chronological age for the two groups of listeners, there was considerable overlap between the range of PPVT ages for the hearing-impaired group and the range of chronological ages for the normal-hearing group.

Speech-Processing Conditions

Four speech processing conditions were examined. These were: (1) original, unmodified speech (U); (2) speech that was uniformly slowed-down or time-expanded by 50% (T); (3) speech modified by 20-dB amplification of time-frequency regions where the critical-band filtered spectral envelope contained energy in the 3-30 Hz range, i.e., amplification of the fast, transitional elements of speech (A); and (4) speech that was time-expanded and had its fast-varying elements amplified (TA).

New recordings of all the speech materials for this study were made by one male talker. This male talker was an experienced speaker, had made recordings for others (including Cochlear

Corporation), and had a typical male fundamental frequency, F_0 (mean F_0 ~ 110 Hz). These recordings served as the unmodified (or unprocessed) speech materials. Speech for the remaining three conditions was processed at Scientific Learning Corporation using the same algorithms employed in their Fast ForWord™ training program. Below is a very brief description of the processing algorithms used in the T, A, and TA conditions. A detailed description is given in Nagarajan et al. (1998). Time-expansion (the T condition) is achieved via a digital signal processing algorithm developed by Portnoff (1981). This algorithm involves computation of the short-time Fourier transform, followed by linear interpolation and phase-modification to a new time-scale, and finally computation of the inverse Fourier transform to yield a time-expanded signal. The time-expansion algorithm is applied uniformly throughout the signal such that all speech segments (formant transitions, steady-state vowels and fricatives, silence gaps, etc.) are lengthened by 50%. For example, a 50-ms formant transition and an 80-ms fricative would become 75-ms and 120-ms in duration, respectively. Envelope amplification (the A condition) is accomplished by an overlap-add procedure. Envelope signals from the equivalent of 22 critical-band-like band-pass filters are found by combining the absolute value of the short-time Fourier transform across the appropriate frequencies for each band signal. These 22 envelope signals are then band-pass filtered (3-30 Hz) and added back to the original envelope signals "to amplify fast-elements while retaining the slower modulations in their original forms" (Nagarajan et al., 1998, p. 261). In addition, a fixed gain is applied to the envelope signals such that the frequency region from roughly 1000-3200 Hz (usually associated with F_2) is amplified by 20 dB. Finally, the entire envelope-modified time signal is obtained by summing the short-time Fourier transforms using a weighted overlap-add procedure. Both the time-expansion (T) and envelope-amplification (A) algorithms were applied to entire original speech signals without further intervention. That is, no phonetic labels or time-markings were used, and no explicit formant manipulations were made. Sample time-waveform and spectrogram displays of the word “bus” are shown in Figure 1 for each of the four speech processing conditions.

Speech Materials

A range of speech materials were selected, from word identification in sentence contexts to CV-syllable contrasts. For each speech-processing condition, the following were employed. First, two lists of revised Bamford-Kowal-Bench (BKB) sentences, consisting of 100 keywords total, were used (Bamford & Wilson, 1979). Second, one list of the Word Intelligibility by Picture Identification (WIPI) test, consisting of 25 words total, was used (Ross & Lerman, 1971). These particular test materials were chosen because they contain vocabulary and syntax appropriate for young children with hearing impairment. Third, eight consonant-vowel (CV) syllables (/da/, /ga/, /ta/, /ti/, /tu/, /sa/, /ʃa/, /za/) were used with the VIDSPAC program (Boothroyd, 1997). Each CV-syllable was represented by three distinct tokens or utterances. These eight CV-syllables were paired to form eight contrasts (/da/ - /ta/, /sa/ - /za/, /da/ - /za/, /sa/ - /ta/, /da/ - /ga/, /sa/ - /ʃa/, /ti/ - /tu/, and /ti/ - /ta), and two contrasts each of the consonant features voicing, manner, and place, as well as two contrasts for vowel identity (height and place).

Presentation and Equipment

The tests were performed inside an IAC sound-isolated booth. The test examiner sat in the IAC booth with the child. For all conditions and for both groups, speech was presented in the free-field using an Anchor AN-100 audio speaker. Free-field presentation, used for all children, was chosen to avoid feedback problems that might occur with the use of headphones on children wearing hearing aids or cochlear implants. Testing was completed in four half-hour sessions for the children with hearing impairment and two one-hour sessions for the children with normal hearing.[2]

All speech was stored digitally with a sampling rate of 22.05 kHz, and each speech waveform was normalized to the same total rms level. The rms-normalization was performed digitally using a custom-written LabView (National Instruments) program. The rms normalization level was chosen such that no digital waveform was clipped when scaled in amplitude. A one-octave band of noise centered at 1 kHz was generated with the same rms level for use as a calibration signal. The sound level at the location of the subject's head for this calibration signal was approximately 74 dBA, measured using a Brüel & Kjær sound-level meter equipped with a #4165 free-field microphone. For unprocessed speech, vowel-peaks in sentences correspond to levels roughly 6 to 12 dB higher than the total rms for a sentence.

A background noise was used for the children with normal hearing. This noise was created to prevent ceiling effects in the scores from this group. The level and spectral shape of this noise was designed to produce elevated audibility thresholds similar to those found for a moderate hearing loss (e.g., those of subject hi8). The noise was generated digitally by filtering white noise through a bank of twenty 1/3-octave, 4th-order Butterworth filters. The spectrum of the background noise is shown in Figure 2. The overall speech-to-noise ratio (SNR) for the children with normal hearing was roughly -4 dB. The background noise was gated on (and off) 50 ms before (and after) the start (and end) of each speech stimulus.

For the BKB and WIPI materials, audio presentation of the speech stimuli was controlled via custom-written LabView programs. Both the sentence and isolated-word speech tests were self-paced, giving the children ample time to respond, and were executed without feedback to the listener. For the BKB sentences, participants were instructed to repeat the sentence that was presented auditorily. Children responded verbally and were encouraged to repeat any word or words they heard. Each sentence was presented only once to each listener, for a total of eight BKB lists (2 BKB lists/child/condition = 32 sentences/child/condition = 100 keywords/child/condition). Responses were generally scored in real-time by the examiner, and were recorded on audiotape for examination at a later time, as needed. Since the equivalent language age for many of the children with impaired hearing was around 5-6 years, and children of that age often make errors in noun-verb agreement, verb tense, etc., the responses to the sentences were scored somewhat liberally. For consistency, this scoring method was applied to all children. A word was scored correct if the root word was perceived correctly. Incorrect word endings, such as “-s” for plurals or “-ed” for verb tense, were ignored.

For the WIPI test, participants were instructed to point to the picture associated with the word that was presented auditorily. The WIPI picture foils were digitally scanned so that responses could be tabulated automatically via a screen-touch or mouse-click. One WIPI list was used per condition (1 list/child/condition = 25 words/child/condition).

The CV-syllable materials were used in a discrimination task that assessed a listener’s ability to hear differences between speech sounds. Syllables were presented via the computer-game-like VIDSPAC program. The VIDSPAC program presents pairs of speech stimuli in a standard-deviant paradigm, in which the standard is presented a random number of times (we chose a uniform distribution between 2 and 5) before the deviant is presented. The listener is instructed to respond when a different syllable sound is heard. For example, for the pair /da/ - /ga/, the first syllable, /da/, is considered the standard and /ga/, the deviant sound. The syllable /da/ might be presented 4 times before /ga/ is presented in the 5th interval. If the listener hears the 5th interval (/ga/) as a sound different from the previous four sounds (in this case, the standard sound

/da/), then the child responds by touching the screen on a designated image or by pressing the spacebar on the keyboard. The listener, in this case, would be given credit for one correct response to the deviant sound. Two types of incorrect response or errors were possible for each “trial” or sequence of standard-deviant sounds. First, if the listener did not detect the deviant sound, i.e., did not make a response when the deviant was presented, then an error of omission was recorded (this reduces the number of “hits”). This type of error is analogous to a “miss” in signal detection theory (Green & Swets, 1974). Second, if the listener incorrectly responds (e.g., by pressing the spacebar) to one of the standard presentations thinking it sounded different from the previous standard presentations, then the VIDSPAC program would record this error as a false positive. This second type of error is analogous to a “false alarm” in signal detection theory (Green & Swets, 1974). The inter-stimulus interval was 1.5 s, and correct/incorrect feedback was provided implicitly through the actions of a cartoon character in the computer-game. Four standard-deviant trials were presented for each CV-syllable pair for each condition. For each presentation interval (standard or deviant), one of the three tokens for each syllable was chosen randomly. Thus, the listener was prevented from responding to either utterance-specific suprasegmental cues (e.g., syllable duration and F_0) or non-phonetic artifacts. VIDSPAC tests were scored automatically by the VIDSPAC computer program.

In each ½-hour session, each hearing-impaired child was randomly assigned (without replacement) two BKB lists, one WIPI list and one CV-list, with the signal processing condition also randomly assigned (without replacement) to each list. Their order of presentation varied randomly from subject to subject. For the children with normal hearing, two equivalent ½-h sessions were combined into one 1-h session.

Results

Listeners with impaired hearing

For the VIDSPAC tests, the reported score is a "corrected-for-chance" score, defined as:

$$"corrected-for-chance"\ score = \frac{\left(h/d\right) - \left(f/s\right)}{\left(1 - f/s\right)} \times 100$$

where

h = number of hits (a response that is a correct detection of the deviant sound),

d = number of deviants presented (deviant trials),

f = number of false positives (incorrect responses to the standard as the deviant), and

s = total number of standards presented.

Figures 3, 4 and 5 show individual data from the BKB, WIPI and VIDSPAC tests, respectively, for the listeners with impaired hearing. For children with impaired hearing, there was considerable variability in individual performance. One likely source of variability was the large variation in severity of hearing loss. For example, listener hi8, with the least severe hearing loss, had roughly the highest overall performance. Variability across individual listeners was greatest for the BKB materials (see Fig. 3) and was greatly diminished for the syllable contrasts (see Fig. 5). Overall, however, the pattern of performance across processing conditions is about the same for each listener.

Figure 6 presents the mean data for the hearing-impaired children. In general, unprocessed speech (U) was the most intelligible condition for words in isolation and in sentences. One notable exception was the performance of listener hi3 who found time-expanded (T) speech most intelligible for all speech materials.

Three repeated-measures ANOVAs were performed on the data in Figure 6, one each for the three types of speech materials (sentences, isolated words, and CV-syllables). For each

ANOVA, there were two factors with two levels each; “time-processing” (levels: *none, time-expansion*) and “amplitude-processing” (levels: *none, envelope-amplification*). For both the sentences and isolated-words, there was a significant effect of “amplitude-processing” on speech intelligibility scores ($F_{1,7} = 39.7$, $p < .001$ and $F_{1,7} = 44.5$, $p < .001$, respectively). For both sentences and isolated-words, there was no effect of “time-processing” on speech intelligibility scores ($F_{1,7} = 4.65$, $p = .07$; $F_{1,7} = 1.18$, $p = .31$), and there was no significant interaction (i.e., “amplitude-processing×time-processing”; $F_{1,7} = .74$, $p = .42$; $F_{1,7} = 1.74$, $p = .23$). Thus, envelope-amplification, with and without time-expansion, degraded the intelligibility of sentences and words, while time-expansion had no effect on intelligibility relative to unprocessed speech. For the overall VIDSPAC results from the CV-syllables, neither time- or amplitude-processing had a statistically significant effect on these subjects’ ability to discriminate syllable pairs ($F_{1,7} = 2.03$, $p = .20$ and $F_{1,7} = 1.06$, $p = .34$). These VIDSPAC results were also analyzed by feature; vowel, voicing, manner and place. There were differences in overall discriminability of these features. In order of increasing difficulty, the corrected-for-chance scores were 95, 85, 78, and 60% for vowel (/i/ vs. /u/, /i/ vs. /a/), manner (/d/ vs. /z/, /s/ vs. /t/), voicing (/d/ vs. /t/, /s/ vs. /z/), and place (/d/ vs. /g/, /s/ vs. /ʃ/) contrasts, respectively. From analogous ANOVA tests, the only significant effect (a negative one) was the effect of amplitude-processing on the discriminability of the manner feature ($F_{1,7} = 18.5$, $p = .004$).

Listeners with normal hearing

Figures 7, 8 and 9 show analogous individual data for the subjects with normal hearing listening in a noise background. For children with normal hearing listening in a noise background, the results were somewhat different. Compared to the data from the hearing-impaired children, there was much less variability in overall performance across these five subjects. This can be expected since these five listeners all had normal hearing and were all subjected to the same SNR during the speech perception tests. The effect of processing on intelligibility is much smaller for listeners with normal hearing than for listeners with impaired hearing, especially for the keywords in sentences (see Fig. 3 vs. Fig. 7).

The mean data for the five normal-hearing children are presented in Figure 10. Again, three repeated-measures ANOVAs were performed on these data, one each for the three types of speech materials. As before, for each ANOVA there were two factors with two levels each; "time-processing" (levels: *none, time-expansion*) and "amplitude-processing" (levels: *none, envelope-amplification*). For both sentences and isolated-words, amplitude-processing had a significant effect on intelligibility ($F_{1,4}$ = 12.0, p = .026 and $F_{1,4}$ = 24.0, p = .008, respectively). For the sentence materials, the effect of time-processing was just significant ($F_{1,4}$ = 8.2, p = .046). All other effects were non-significant. Specifically, the interaction of time-processing×amplitude-processing was not significant for both sentences and words ($F_{1,4}$ = 3.08, p = .15; $F_{1,4}$ = 3.58, p = .13), and time-processing was not significant for isolated-words ($F_{1,4}$ = 6.59, p = .062). Thus, for normal-hearing listeners, envelope-amplification had a degrading effect on the intelligibility of words and sentences relative to unprocessed speech. For CV-syllables, neither time-processing nor amplitude-processing had a statistically significant effect on the ability to discriminate syllable pairs ($F_{1,4}$ = 1.01, p = .37 and $F_{1,4}$ = .002, p = .96, respectively), and there was no interaction ($F_{1,4}$ = 4.27, p = .11). For these CV-syllables, overall discriminability of features was easiest for the voicing contrast, followed by vowel, manner and place, with corrected-for-chance scores of 94, 92, 89, and 71%, respectively. Due to a large number (half or more) of perfect scores (100% correct discrimination) for the voicing, vowel and manner features, these features were not subjected to further analyses. For the remaining feature, place, time-processing had a significant degrading effect on its discrimination ($F_{1,4} = 10.0$, $p = .034$).

Discussion

The pattern of results from the two listener groups in this study, children with and without impaired hearing, is fairly similar. For both listener groups and all types of speech materials, neither time-expansion nor envelope-amplification provided an advantage in speech intelligibility relative to unprocessed speech. Also, for both listener groups there is no effect on CV-syllable discrimination performance due to speech processing. For this task, it is certainly imaginable that processing could have either increased or decreased discrimination ability by making the syllable pairs more or less distinct from each other. Yet, for both these listener groups neither time-expansion nor envelope-amplification had a statistically significant effect on overall syllable discrimination. Though syllable-discrimination performance is fairly good in all processing conditions for both groups of listeners, we cannot infer how the processed speech segments might be labeled. That is, we cannot say, for example, whether an A-processed /ga/ would be recognized or labeled as /ga/. The pattern of results for feature discrimination of these CV-syllables is also fairly similar for the two listener groups. For both groups, vowel and manner contrasts were easily discriminated, and place discrimination was most difficult. This result is consistent with many other studies that found place perception to be very difficult (e.g., Carney et al., 1993; Miller & Nicely, 1955; Tyler, 1990). However, our two listener groups differed in their ability to discriminate CV-syllables that varied only in their voicing feature (/sa/ vs. /za/ and /da/ vs. /ta/). The normal-hearing listeners had little problem with voicing discrimination (94% correct) while the hearing-impaired listeners had more difficulty (78% correct).

The listener groups differ somewhat in the exact pattern of processing effects on intelligibility. For the children with hearing impairment, envelope-amplification degraded intelligibility for both words in sentences and words in isolation. This degradation occurred when the envelope-amplification was applied by itself (A) or in combination with time-expansion (TA). For isolated-words, the normal-hearing listeners exhibited the same pattern of results in that envelope-amplification, alone or combined with time-expansion, degraded intelligibility relative to unprocessed speech. However, for sentence materials, both time- and amplitude-processing degraded intelligibility for the normal hearing listeners. We offer no explanation for this particular difference between the two subject groups.

The size of the degradation effect due to envelope-amplification is also different for the two subject groups, and seems to depend on the type of speech material employed. For the hearing-impaired children, relative to unprocessed speech, envelope-amplification by itself reduced intelligibility scores by 21 (from 69 to 48%) and 30 (from 67 to 37%) percentage points for words and sentences, respectively. The analogous reductions for the children with normal hearing were 18 (from 72 to 54%) and 8 (from 90 to 82%) percentage points. Thus, compared to children with normal hearing, the hearing-impaired children exhibited larger degradations for envelope-amplified speech relative to unprocessed speech and an opposite effect of sentence context. That is, for normal-hearing listeners, there is a smaller degradation due to A-processing for sentence materials than for isolated words. However, for the hearing-impaired listeners the opposite is found: there is a larger degradation due to A-processing for sentence materials than for isolated words. Children with normal hearing, who have more developed language skills, may be better able to take advantage of linguistic context to overcome speech degradations when listening to sentence materials. This view is supported by the approximately equal intelligibility of sentences and words found for the hearing-impaired listeners (67 and 69%, respectively) as compared to the greater intelligibility of sentence materials relative to isolated words for the normal-hearing listeners (90 and 72%, respectively). These results indicate that sentence context is probably not being utilized by the hearing-impaired children.

Our results with time-expanded speech (uniform time-expansion by 50% for all sounds) are generally consistent with previously reported findings (Schon, 1970; Picheny et al., 1989; Uchanski et al., 1996). For both listener groups and all speech materials, the perception of time-expanded speech was not significantly different from unprocessed speech – except for a slight degrading effect for normal-hearing children listening to BKB sentences. Thus, for listeners like ours, there is no evidence that time-expansion is beneficial for speech intelligibility.

For the envelope-amplification processing, a comparison of our intelligibility results with data from others is more difficult. The method of envelope-amplification used in this study consists of two sub-components: i) band-pass filtering of the envelope (or modulation) spectrum in the 3-30 Hz region; and ii) a gain of 20 dB for the analysis bands in the filter-bank with center frequencies between 2000 and 4000 Hz. We did not evaluate these two sub-components separately for their effects on intelligibility. While band-pass filtering of the envelope spectrum has not been studied, low-pass and high-pass filtering of the envelope spectrum have been investigated (Drullman, Festen & Plomp, 1994a; Drullman, Festen & Plomp, 1994b). In Drullman et al. (1994a), low-pass filtering of the modulation spectrum degraded speech intelligibility when the cutoff frequency was less than or equal to 16 Hz. Analogously, high-pass filtering of the modulation spectrum degraded speech intelligibility when the cutoff frequency was greater than or equal to 8 Hz (Drullman, Festen & Plomp, 1994b). Thus, it might appear reasonable to assume that band-pass filtering (3 to 30 Hz) of the modulation spectrum would have no detrimental effect on intelligibility. However, band-pass filtering of the modulation spectrum has not been examined explicitly and, due to redundancies in the speech signal, combining low-pass and high-pass filtering results could be misleading.

Finally, while the data in this study show no intelligibility benefit from either time-expansion or envelope-amplification, these results are not necessarily in conflict with those of Tallal et al. (1996). There were many important differences between this study and theirs. First, we employed hearing-impaired and normal hearing listeners (in noise), not language-impaired listeners. Second, we were interested in speech intelligibility as the outcome measure whereas speech intelligibility was never of concern in the design or development of their speech modification algorithm. Third, we did not train our listeners extensively with the processed speech materials as was done in the Tallal et al. study.

References

Alexander, D. W. & Frost, B. P. (1982). Decelerated synthesized speech as a means of shaping speed of auditory processing of children with delayed language. *Perceptual & Motor Skills*, *55*(3 Pt 1), 783-792.

Bamford, J. & Wilson, I. (1979). Methodological Considerations and Practical Aspects of the BKB Sentence Lists. *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children*. J. Bench and J. Bamford. London: Academic Press**,** 147-187.

Boothroyd, A. (1997). VIDSPAC 2.0: A video game for assessing speech pattern contrast perception in children. San Diego.

Boothroyd, A., Geers, A., & Moog, J. (1991). Practical Implications of Cochlear Implants in Children. *Ear and Hearing*, *12*(4), 81S-89S.

Bradlow, A. R., Kraus, N., Nicol, T., McGee, T., Cunningham, J., Zecker, S., & Carrell, T. (1999). Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. *The Journal of the Acoustical Society of America*, *106*(4), 2086-2096.

Carney, A., Osberger, M., Carney, E., Robbins, A., Renshaw, J., & Miyamoto, R. (1993). A comparison of speech discrimination with cochlear implants and tactile aids, *The Journal of the Acoustical Society of America, 94*(4), 2036-2049.

Drullman, R., Festen, J. M., & Plomp, R. (1994a). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America, 95*(2), 1053-1064.

Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, *95*(5), 2670-2680.

Dunn, L. & Dunn, L. (1981). Peabody Picture Vocabulary Test - Revised. Circle Pines, MN: American Guidance.

Fosler-Lussier, E. & Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication, 29*, 137-158.

Fryauf-Bertschy, H., Tyler, R. S., Kelsay, D. M. R., Gantz, B. J., & Woodworth, G. G. (1997). Cochlear Implant Use by Prelingually Deafened Children: The Influences of Age at Implant and Length of Device Use. *Journal of Speech, Language, and Hearing Research, 40*, 183-199.

Geers, A. E. & Moog, J. (1992). Speech perception and production skills of students with impaired hearing from oral and total communication education settings. *Journal of Speech and Hearing Research, 35*, 1384-1393.

Gordon-Salant, S. (1986). Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing. *The Journal of the Acoustical Society of America, 80*(6), 1599-1607.

Green, D., & Swets, J. (1974). *Signal Detection Theory and Psychophysics,* Robert E. Krieger Publishing Co., Inc., Huntington, NY.

Hazan, V. & Simpson, A. (1998). The Effect of Cue-Enhancement on the Intelligibility of Nonsense Word and Sentence Materials Presented in Noise. *Speech Communication 24,* 211-226.

Kirk, K. I., Pisoni, D. B., & Osberger, M. J. (1995). Lexical Effects on Spoken Word Recognition by Pediatric Cochlear Implant Users. *Ear & Hearing, 16*(5), 470-481.

Korabic, E. W., Freeman, B. A., & Church, G. T. (1978). Intelligibility of time-expanded speech with normally hearing and elderly subjects. *Audiology, 17*(2), 159-164.

Krause, J. (1995). *The Effects of Speaking Rate and Speaking Mode on Intelligibility*. unpublished doctoral thesis, Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

McAnally, K. I., Hansen, P. C., Cornelissen, P. L., & Stein, J. F. (1997). Effect of Time and Frequency Manipulation on Syllable Perception in Developmental Dyslexics. *Journal of Speech, Language, and Hearing Research, 40,* 912-924.

Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal Processing Deficits of Language-Learning Impaired Children Ameliorated by Training. *Science, 271*, 77-81.

Merzenich, M., Miller, S., Jenkins, W., Saunders, G., Protopapas, A., Peterson, B. & Tallal P. (1998). Amelioration of the acoustic and speech reception deficits underlying language-based learning impairments. In C. von Euler (Ed), *Basic Mechanisms in Cognition and Language*, (pp. 143-172). Amsterdam: Elsevier.

Miller, G., & Nicely, P. (1955). An Analysis of Perceptual Confusions among Some English Consonants. *The Journal of the Acoustical Society of America, 27*(2), 338-352.

Moon, S.-J. & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America*, *96*, 40-55.

Moore, B. C. J., Peters, R. W., & Stone, M. A. (1999). Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. *The Journal of the Acoustical Society of America, 105*(1), 400-411.

Moore, K. C. & Zue, V. W. (1985). The effect of speech rate on the application of low-level phonological rules in American English. *The Journal of the Acoustical Society of America, 77*(Suppl. 1), S53.

Nagarajan, S. S., Wang, X., Merzenich, M. M., Schreiner, C. E., Johnston, P., Jenkins, W. M., Miller, S., & Tallal, P. (1998). Speech Modification Algorithms Used for Training Language Learning-Impaired Children. *IEEE Transactions on Rehabilitation Engineering, 6*(3), 257-268.

Payton, K. L., Uchanski, R. M., & Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America, 95*(3), 1581-1592.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking Clearly for the Hard of Hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28,* 96-103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking Clearly for the Hard of Hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29,* 434-446.

Picheny, M. A., Durlach, N. I., & Braida, 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. *Journal of Speech and Hearing Research, 32,* 600-603.

Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *The Journal of the Acoustical Society of America, 83*(6), 2322-2327.

Portnoff, M. R. (1981). Time scale modification of speech based on short-time Fourier Analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-29,* 374-390.

Ross, M. & Lerman, J. (1971). Word Intelligibility by Picture Identification. Pittsburgh: Stanwix House.

Schmitt, J. F. (1983). The effects of time compression and time expansion on passage comprehension by elderly listeners. *Journal of Speech and Hearing Research, 26*, 373-377.

Schon, T. D. (1970). The effects on speech intelligibility of time-compression and -expansion on normal-hearing, hard of hearing, and aged males. *Journal of Auditory Research, 10*, 263-268.

Stollman, M. H. P., Kapteyn, T. S., & Sleeswijk, B. W. (1994). Effect of time-scale modification of speech on the speech recognition threshold in noise for hearing-impaired and language-impaired children. *Scandinavian Audiology, 23*(1), 39-46.

Sussman, J. E. & Carney, A. E. (1989). Effects of transition length on the perception of stop consonants by children and adults. *Journal of Speech and Hearing Research, 32*, 151-160.

Tallal, P. & Piercy, M. (1975). Developmental Aphasia: The Perception of Brief Vowels and Extended Stop Consonants. *Neuropsychologia, 13*, 69-74.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., & Merzenich, M. M. (1996). Language Comprehension in Language-Learning Impaired Children Improved with Acoustically Modified Speech. *Science, 271*, 81-84.

Tyler, R. (1990). What Should be Implemented in Future Cochlear Implants? *Acta Otolaryngology*, Suppl. 469, 268-275.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking Clearly for the Hard of Hearing IV: Further Studies of the Role of Speaking Rate. *Journal of Speech and Hearing Research, 39*, 494-509.

Author Note

Athanassios Protopapas, Institute for Language and Speech Processing, Department of Educational Technology, Athens, Greece.

This work was supported by an anonymous, private foundation. The authors acknowledge and thank Scientific Learning Corporation for its cooperation and generosity in processing the original speech materials.

Contact author: Rosalie M. Uchanski, Central Institute for the Deaf, 4560 Clayton Avenue, Saint Louis, MO 63110, USA, or ruchanski@cid.wustl.edu.

Footnotes

[1] The speech enhancements generated by Hazan and Simpson rely on manual selection and segmentation, and presently cannot be automated.

[2] The schedule of sessions was different for normal-hearing (NH) and hearing-impaired children (HI) because of a) time constraints within CID’s school day (for the HI children) and b) a desire to minimize, within reason, the number of trips made to CID (for the NH children). A break was given to the NH children at the halfway point of their 1-hour sessions.

Table 1. Characteristics of children with hearing impairment who participated in this study. Unaided pure-tone-average (PTA) is from 500, 1000 and 2000 Hz. "ci" represents a Nucleus cochlear implant and "ha" represents a hearing aid. PPVT score represents the Peabody Picture Vocabulary Test result in equivalent language age.

	PTA Right, Left	Device(s)		Age	PPVT score
Subject No.	(dB HL)	Type	Ear(s)	(years: months)	(years: months)
hi1	106, 108	ci	L	11:5	5:1
hi2	120, 120	ci	R	8:10	5:6
hi3	116, 116	ci/ci	R/L	13:10	5:2
hi4	116, 111	ci	L	11:5	5:0
hi5	98, 98	ha	L	14:9	6:2
hi6	75, 75	ha	R	13:10	10:9
hi7	116, 106	ci/ha	R/L	11:9	6:7
hi8	38, 36	ha/ha	R/L	12:11	9:4

Figure Captions

Figure 1. Time waveforms and spectrograms of the word “bus” for each of the four processing conditions. “U” represents unprocessed speech, “T” represents time-expanded speech, “A” represents envelope-amplified speech, and “TA” represents speech that is both time-expanded and envelope-amplified.

Figure 2. Spectrum of background noise used for listeners with normal hearing.

Figure 3. Percent correct keyword scores for BKB sentences presented to hearing-impaired listeners. “U” represents unprocessed speech, “T” represents time-expanded speech, “A” represents envelope-amplified speech, and “TA” represents speech that is both time-expanded and envelope-amplified. Data from individual listeners are shown for each speech-processing condition. The percent correct score is from 100 keywords per listener per condition.

Figure 4. Percent correct score for WIPI lists presented to hearing-impaired listeners. “U” represents unprocessed speech, “T” represents time-expanded speech, “A” represents envelope-amplified speech, and “TA” represents speech that is both time-expanded and envelope-amplified. Data from individual listeners are shown for each speech-processing condition. The percent correct score is from 25 words (1 WIPI list) per listener per condition.

Figure 5. Overall syllable discrimination score for syllable-pairs presented to hearing-impaired listeners. “U” represents unprocessed speech, “T” represents time-expanded speech, “A” represents envelope-amplified speech, and “TA” represents speech that is both time-expanded and envelope-amplified. Data from individual listeners are shown for each speech-processing condition. The score reported is corrected-for-chance performance.

Figure 6. Summary of intelligibility and discrimination results for hearing-impaired listeners. “U” represents unprocessed speech, “T” represents time-expanded speech, “A” represents envelope-amplified speech, and “TA” represents speech that is both time-expanded and

envelope-amplified. For each type of speech material tested, the average performance and +/- one standard deviation across subjects are shown.

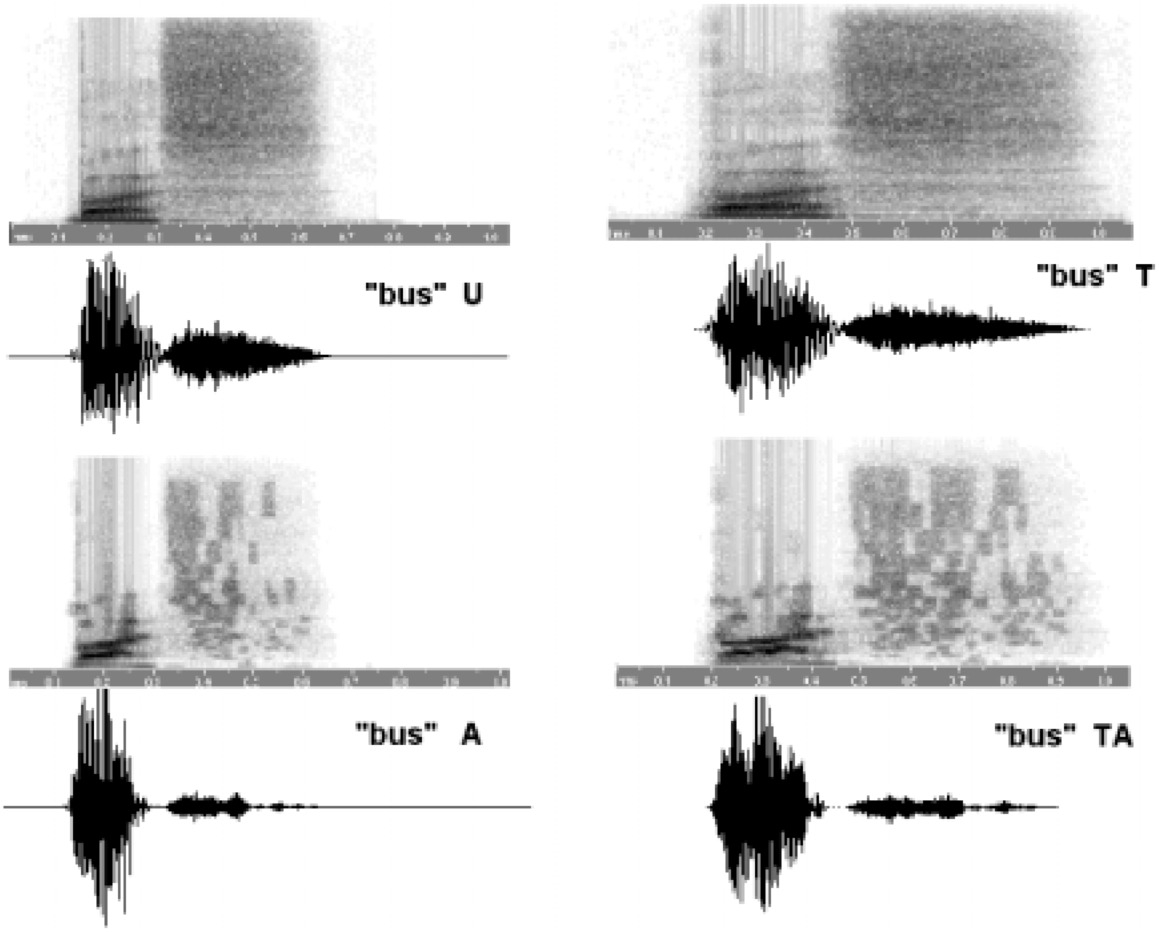
Figure 7. Percent correct keyword score for BKB sentences presented to normal hearing subjects listening in noise. “U” represents unprocessed speech, “T” represents time-expanded speech, “A” represents envelope-amplified speech, and “TA” represents speech that is both time-expanded and envelope-amplified. Data from individual listeners are shown for each speech-processing condition. The percent correct score is from 100 keywords per listener per condition.

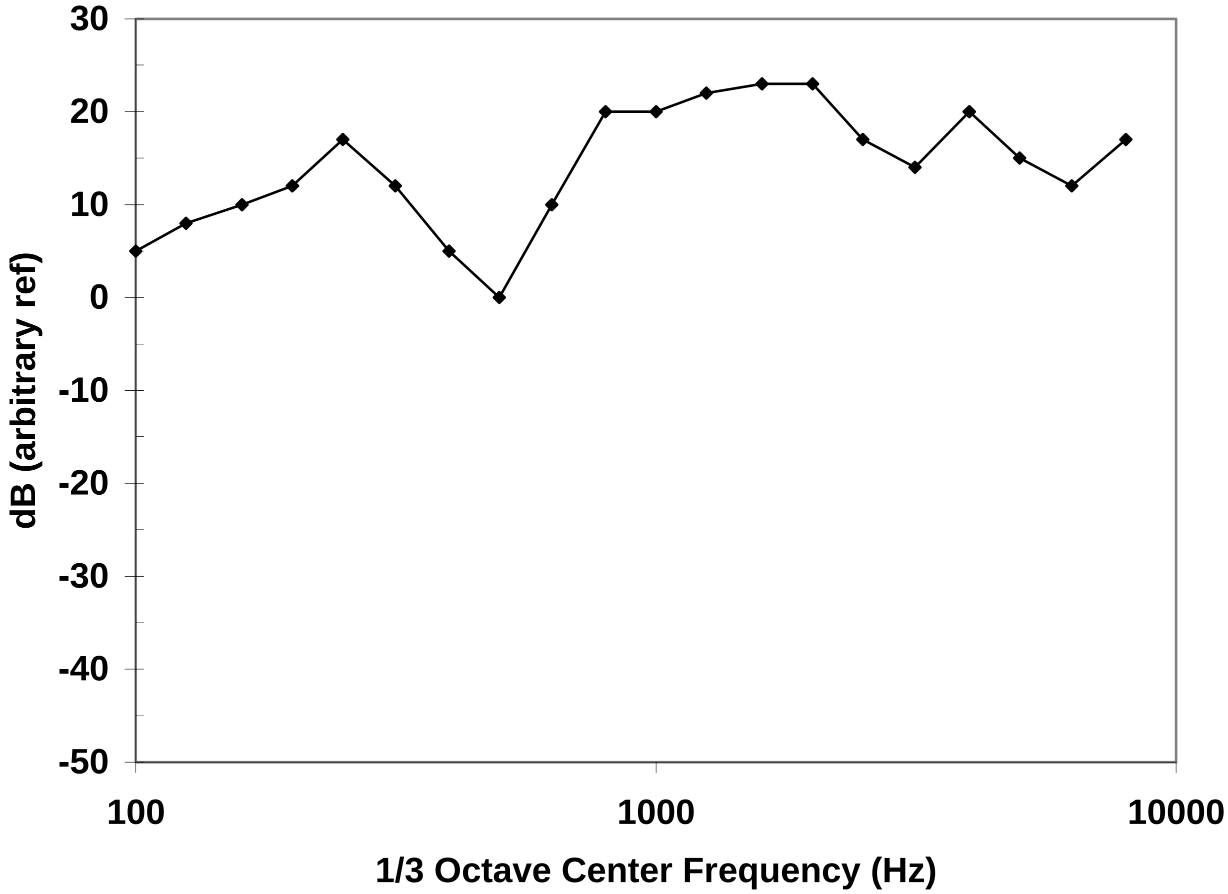
Figure 8. Percent correct word score for WIPI lists presented to normal hearing subject listening in noise. “U” represents unprocessed speech, “T” represents time-expanded speech, “A” represents envelope-amplified speech, and “TA” represents speech that is both time-expanded and envelope-amplified. Data from individual listeners are shown for each speech-processing condition. The percent correct score is from 25 words (1 WIPI list) per listener per condition.

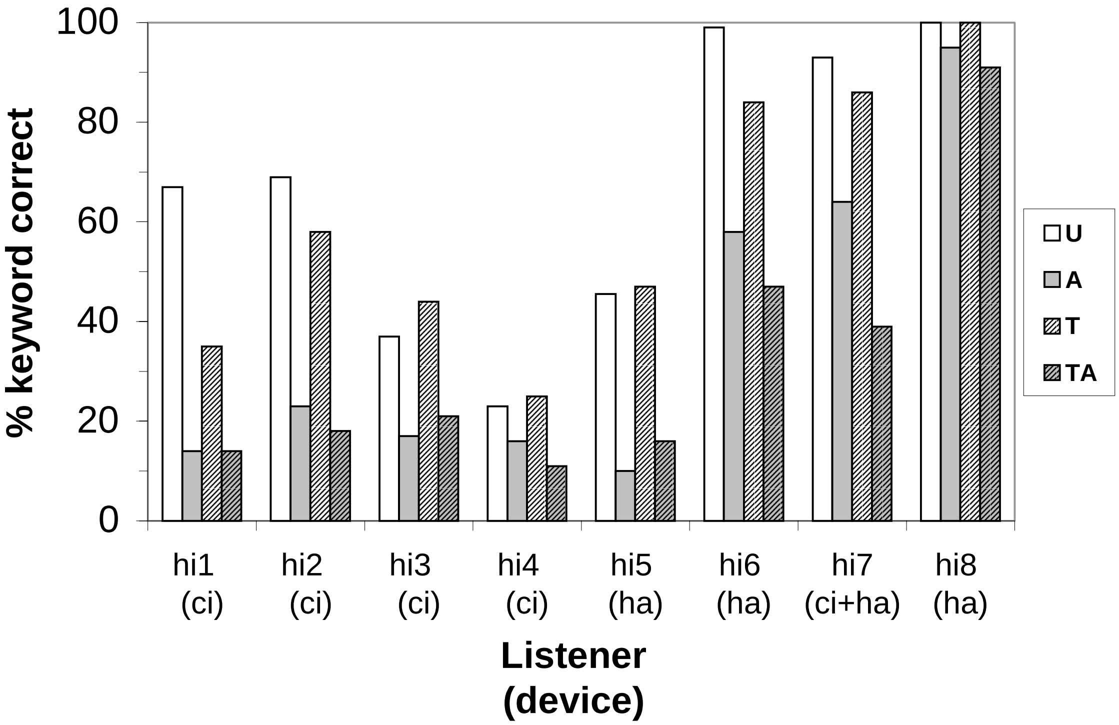
Figure 9. Overall syllable discrimination score for syllable-pairs presented to normal hearing subjects listening in noise. “U” represents unprocessed speech, “T” represents time-expanded speech, “A” represents envelope-amplified speech, and “TA” represents speech that is both time-expanded and envelope-amplified. Data from individual listeners are shown for each speech-processing condition. The score reported is corrected-for-chance performance.

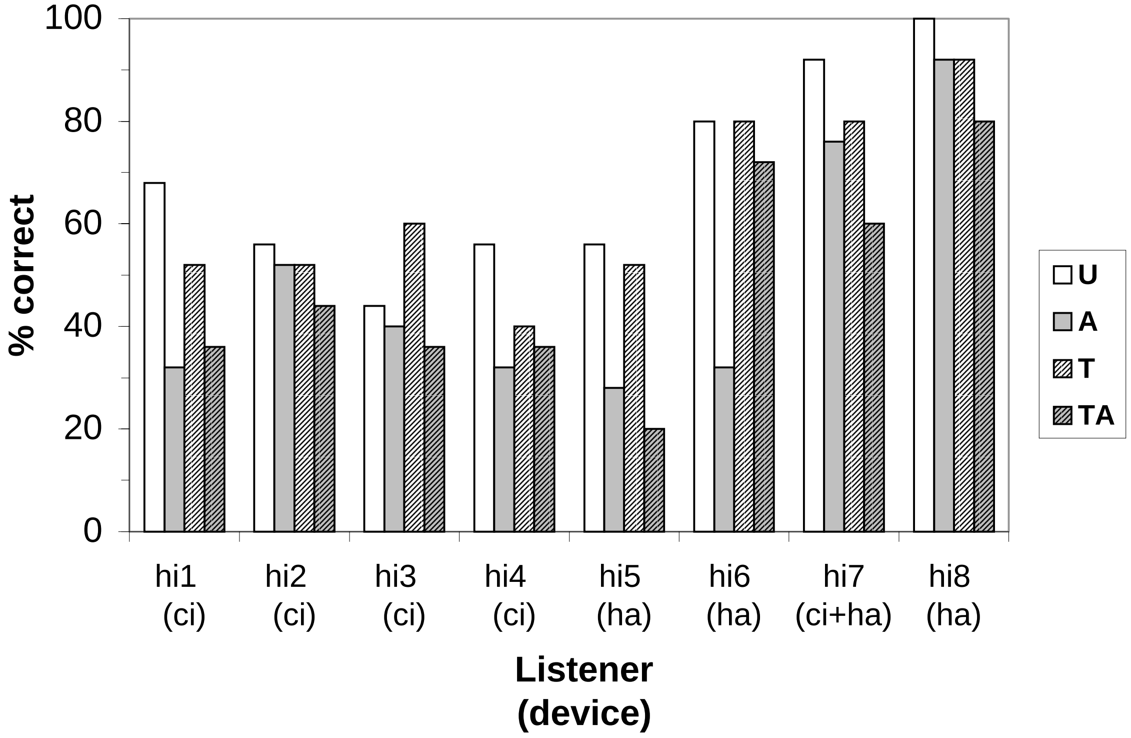
Figure 10. Summary of intelligibility and discrimination results for normal hearing subjects listening in noise. “U” represents unprocessed speech, “T” represents time-expanded speech, “A” represents envelope-amplified speech, and “TA” represents speech that is both time-expanded and envelope-amplified. For each type of speech material tested, the average performance and +/- one standard deviation across subjects are shown.

FIGURE 1

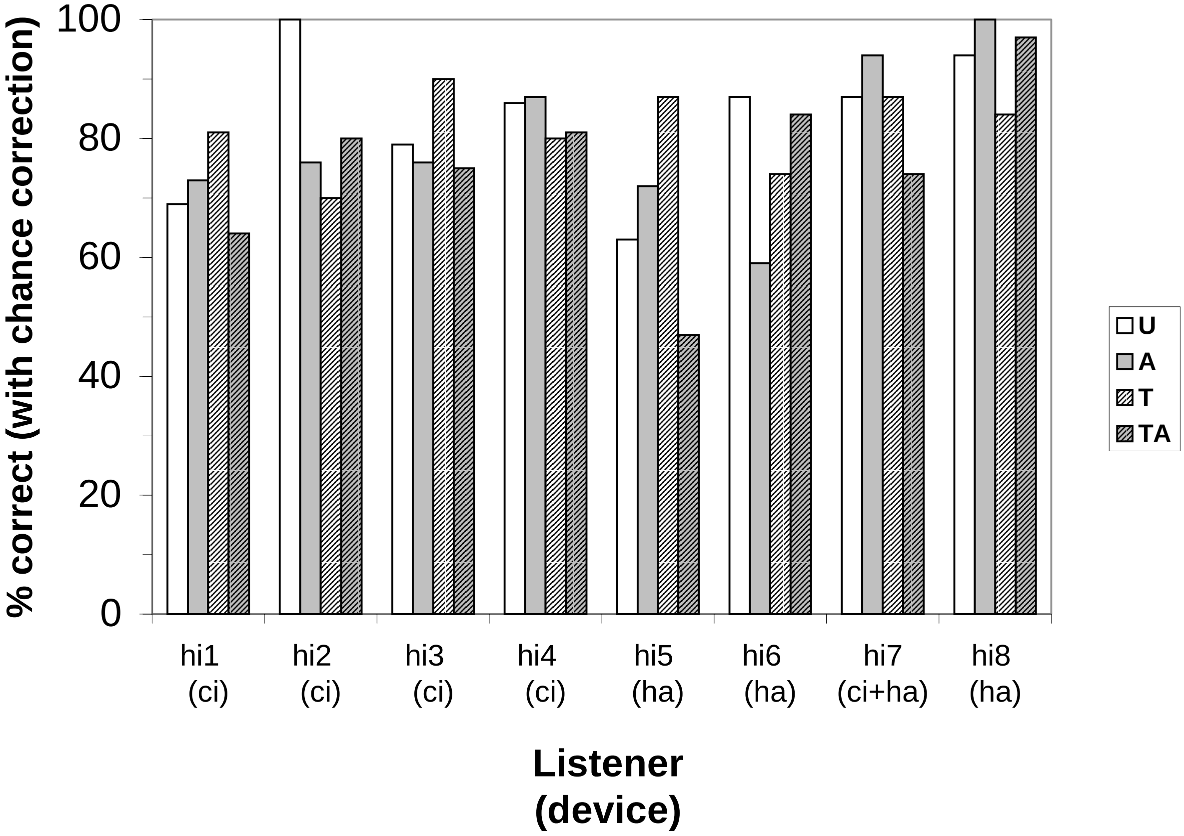


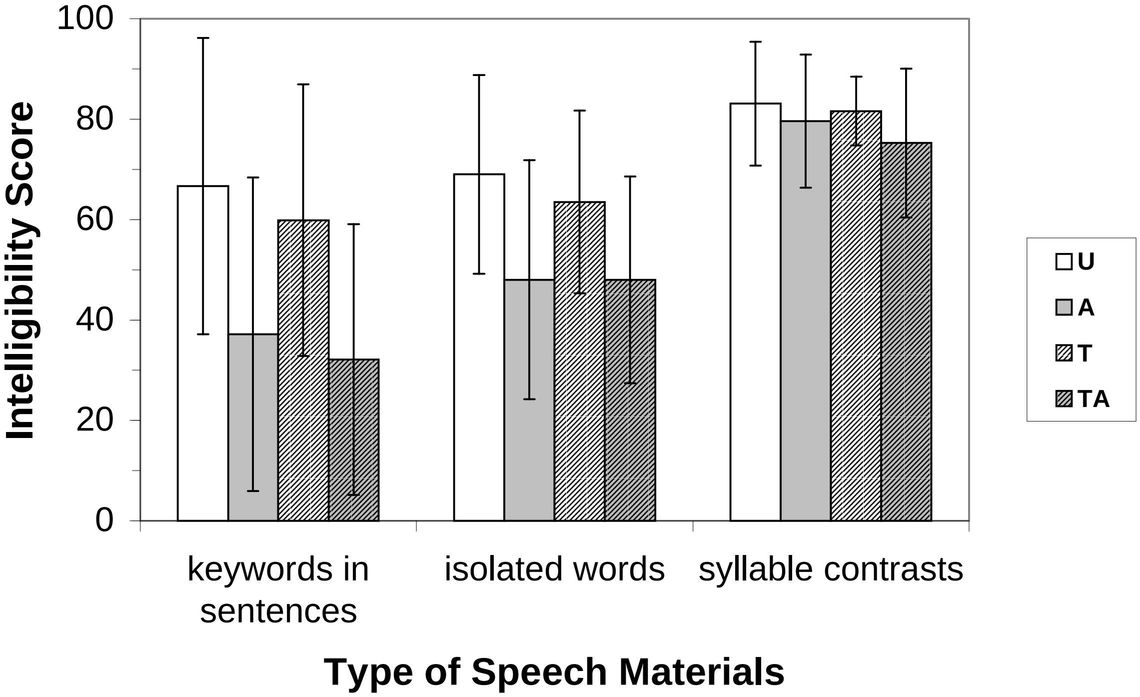

30
20
10
0
-10
-20
-30
-40
-50
dB (arbitrary ref)
100
1000
10000
1/3 Octave Center Frequency (Hz)

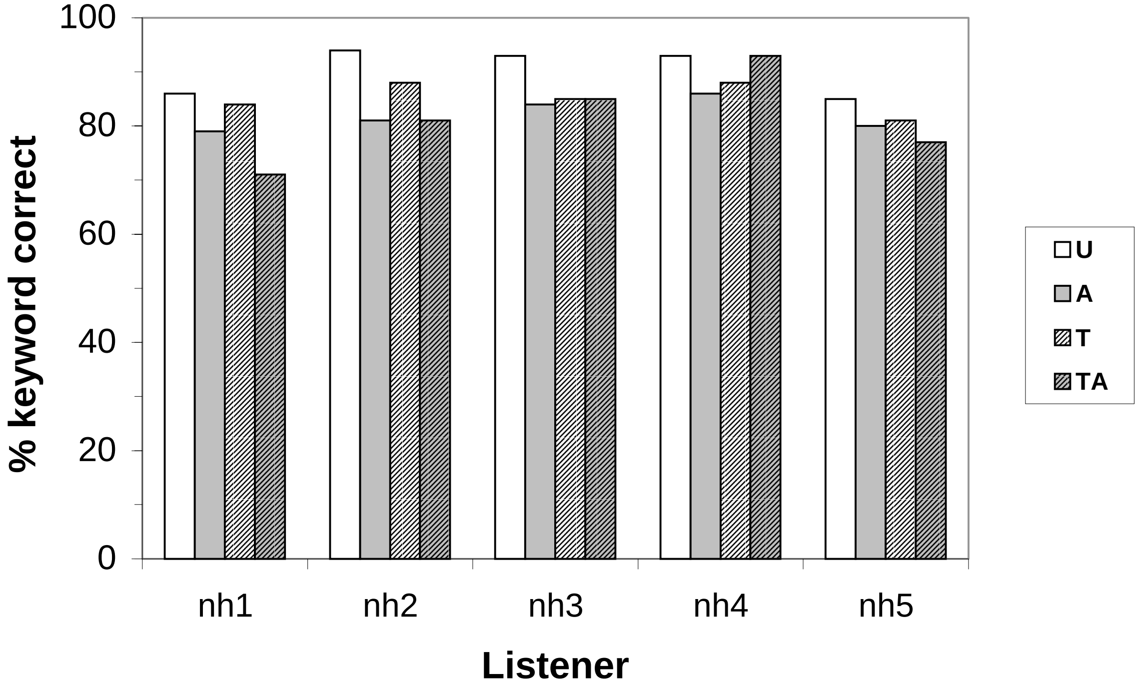

% keyword correct
100
80
60
40
20
0
hi1 (ci)
hi2 (ci)
hi3 (ci)
hi4 (ci)
hi5 (ha)
hi6 (ha)
hi7 (ci+ha)
hi8 (ha)
Listener (device)
U
A
T
TA

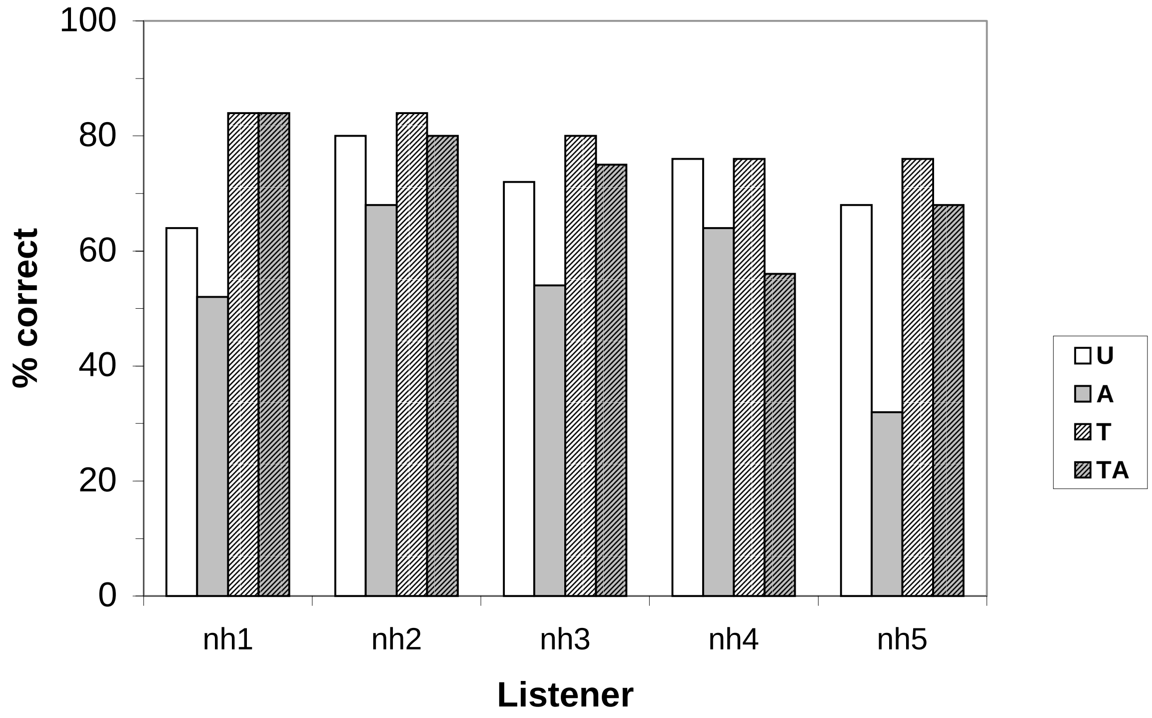


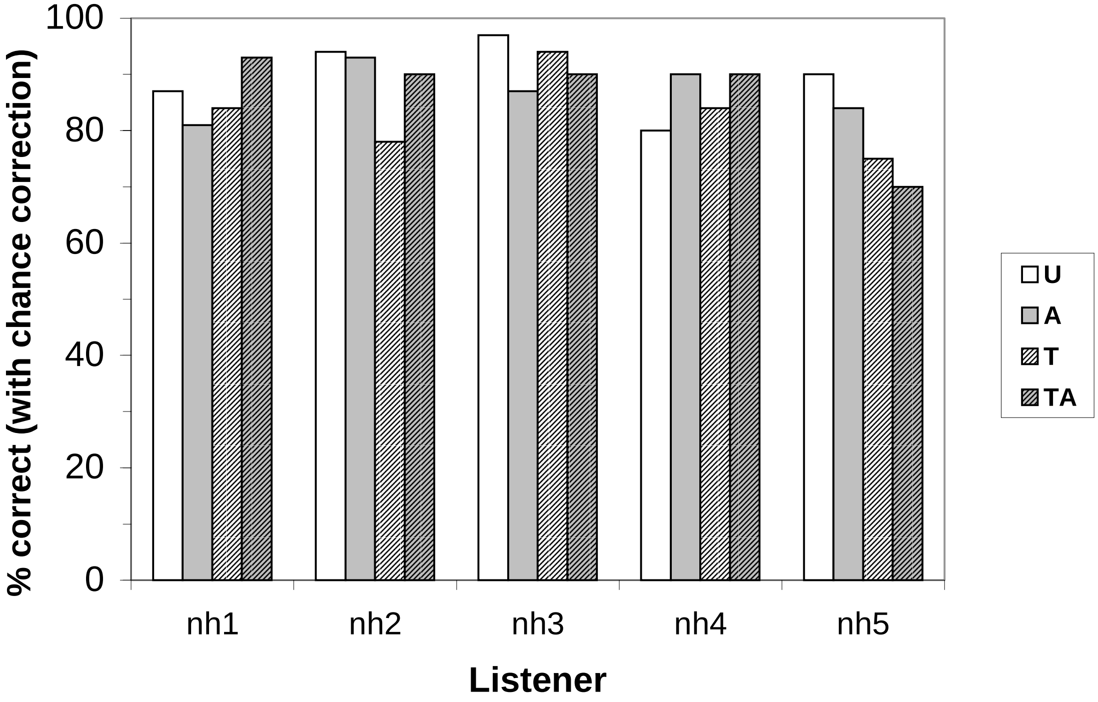
100
80
60
40
20
0
% correct
hi1 (ci)
hi2 (ci)
hi3 (ci)
hi4 (ci)
hi5 (ha)
hi6 (ha)
hi7 (ci+ha)
hi8 (ha)
Listener (device)
U
A
T
TA

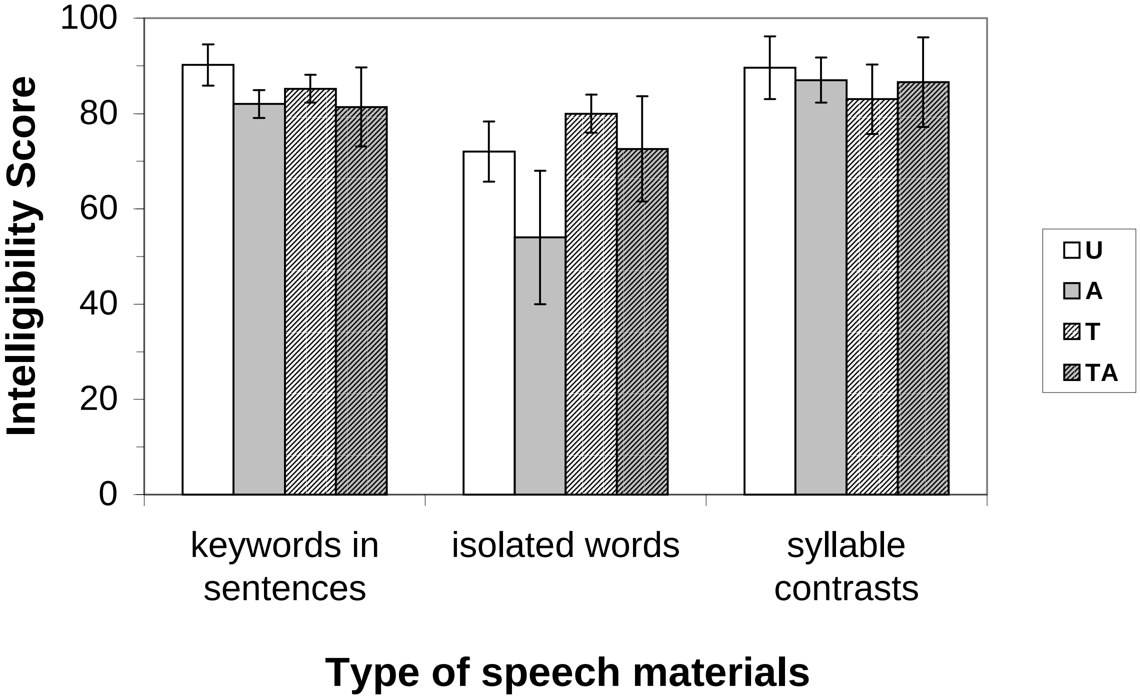

% correct (with chance correction)
100
80
60
40
20
0
hi1 (ci)
hi2 (ci)
hi3 (ci)
hi4 (ci)
hi5 (ha)
hi6 (ha)
hi7 (ci+ha)
hi8 (ha)
Listener (device)
U
A
T
TA


100
80
60
40
20
0
Intelligibility Score
keywords in sentences
isolated words
syllable contrasts
Type of Speech Materials
U
A
T
TA


100
80
60
40
20
0
% keyword correct
nh1
nh2
nh3
nh4
nh5
Listener
U
A
T
TA


100
80
60
40
20
0
% correct
nh1
nh2
nh3
nh4
nh5
Listener
U
A
T
TA


100
80
60
40
20
0
% correct (with chance correction)
nh1
nh2
nh3
nh4
nh5
Listener
U
A
T
TA


Intelligibility Score
100
80
60
40
20
0
keywords in sentences
isolated words
syllable contrasts
Type of speech materials
□ U
▨ A
▨ T
▨ TA