CONTRIBUTION OF CONSONANT VERSUS VOWEL INFORMATION TO SENTENCE INTELLIGIBILITY BY NORMAL AND HEARING-IMPAIRED LISTENERS

T. Zachary Burkle

Accepted by the Faculty of the Graduate School, Indiana University, in partial fulfillment of the requirements for Master of Arts degree in the Department of Speech and Hearing Sciences.

Diane Kewley-Port, Ph.D.

Masters Thesis
Committee

Larry E. Humes, Ph.D.

Jennifer J. Lentz, Ph.D.

Date

ABSTRACT

T. Zachary Burkle

CONTRIBUTION OF CONSONANT VERSUS VOWEL INFORMATION TO SENTENCE INTELLIGIBILITY BY NORMAL AND HEARING-IMPAIRED LISTENERS

The purpose of this study was to examine the contribution of information provided by vowels versus consonants to sentence intelligibility in young normal-hearing (YNH) and elderly hearing-impaired (EHI) listeners. Sentences were presented in three conditions, with either the vowels or the consonants replaced with speech shaped noise, or unaltered. Sentences from male and female talkers in the TIMIT database were selected. EHI subjects listened at 95 dB SPL, and YNH subjects at both 95 and 70 dB SPL. Subjects listened to each sentence twice and were asked to repeat the entire sentence after each presentation. Words were scored correct if identified exactly. Average performance for unaltered sentences was greater than 94%. Vowel-present conditions were always significantly more intelligible than consonant-present conditions, similar to data reported by Cole and colleagues [Proceedings of ICASSP, 1996]. Across groups, performance in the vowel-present conditions exceeded that in the consonant-present conditions by 14 to 40%, although EHI subjects performed more poorly than YNH subjects. In contrast to written English, vowels in spoken language carry more information about sentences than consonants for both normal and hearing-impaired listeners.

TABLE OF CONTENTS

CHAPTER Page

Chapter I. Introduction and Overview of Literature

It has become almost a commonplace statement in intelligibility testing that most of the information in speech is carried by the consonants (Owens and Talbott, 1968). Verification of this statement can be found in research carried out by Fletcher (1929). Those experiments focused on “letter articulation”, “consonant articulation”, and “vowel articulation” within CV (consonant-vowel), VC (vowel-consonant), and CVC (consonant-vowel-consonant) syllables over a transmission system in which the frequency bandwidth was altered. Observers wrote the sounds in that they heard. The correct responses and errors for each of the fundamental sounds were recorded and analyzed as articulation scores. The letter articulation, the percent of correct letter sounds, was 72.2 percent correct. Consonant articulation was 65.8 percent correct and vowel articulation was 83.4 percent correct. Because the consonants were usually harder to recognize, it was concluded that if consonants were correctly identified, they would contribute more towards the overall intelligibility than vowels. While this argument is a logical conclusion given the results, note that this experiment, and others like it, did not manipulate vowel and consonant sounds directly but rather the properties of the transmission channel.

Very few researchers have examined the contributions of vowel and consonantal information separately to spoken English, and even fewer have examined how cochlear hearing impairment may affect these contributions. It is obvious that for a person with a high frequency sloping audiogram, the audibility of higher-frequency consonant energy is sacrificed, often being completely inaudible. Once audibility is ensured, however, most researchers would still hypothesize that consonants carry most of the crucial information required for speech intelligibility. Miller (1951) states that the weakness in intensity of the consonant sounds is unfortunate, because the consonants are more critical for the correct interpretation of speech. The present study examines these contributions of vowels and consonants to sentence intelligibility in both young normal-hearing and in elderly hearing-impaired participants.

Speech communication is so important that it is rightly considered to be the most characteristic feature of the human race (Plomp, 2002). The acoustic information carried by speech is quite complex and has many dynamic variations. Sounds are by their nature dynamic, changing over time in terms of level and spectral content. In general, it has been observed that consonant information has predominantly high frequency information above 2000 Hz, contrasting with vowels which contain predominantly mid to low frequencies below 2000 Hz. Numerous studies have looked at the variation of the frequency spectrum over a finite period of time. Among the earliest discoveries in speech perception research was the importance of formant transitions for the perception of place of articulation in syllable-initial and syllable-final stop consonants (Cooper et al, 1952). It has been shown by formant analysis that the dynamic spectral variation in vowels provides reliable acoustic cues in fluent speech that contribute to both consonant and vowel identification (Liberman et al, 1967). Therefore it appears that information specifying consonants may be more broadly distributed across the frequency spectrum while vowel information is more concentrated in the lower frequencies.

The process of following rapid changes, and integrating information, over time is referred to as temporal processing (Oxenham and Bacon, 2003). However, several streams of information in the speech signal with different rates appear to contribute to it's complex temporal structure. Plomp (1983) observed that the speech waveform is temporally modulated at two different rates; at the more rapid fundamental frequency of vocal fold vibration, and at the slower rate of vocal tract opening/closing (Van Tasell, 1993). In addition, there appear to be two different sets of gestures with the rapid articulations for consonants superimposed on the slower ones for vowels. Since the frequency content of speech changes over time, it is these changes that carry information about the vocal tract gestures that produce speech sounds, and therefore about the phonemic identities of the individual speech sounds (Van Tasell, 1993). Temporal speech information is likely to have special significance for hearing-impaired listeners, for users of cochlear implants, and more generally, in listening conditions where background noise masks the spectral structure (Faulkner and Rosen, 1999). This suggests that temporal cues and changes of gross spectral shape may be of greater significance in speech perception than has often been thought (Faulkner and Rosen, 1999). Poor temporal resolution may include a reduced ability to take advantage of momentary dips in the level of background sounds to hear out the sounds of interest. This may be particularly important in real-world situations, where background noises are often not continuous, steady-state sounds, but fluctuate from moment to moment (Oxenham and Bacon, 2003).

The speech signal is also highly variable in terms of its intensity. Research on speech acoustics has shown that the intensity of most consonant sounds is substantially lower than that of vowels, with the weakest consonant being as much as 30 dB lower than the strongest vowels (Dunn and White, 1940). Several investigators have studied the relationship in consonant and vowel amplitude ratios to speech intelligibility. Gordon-Salant (1986) showed that increasing consonant to vowel ratios by 10 dB improved speech recognition for consonant-vowel (CV) nonsense stimuli for both young and elderly normal-hearing listeners at two signal levels (75 and 90 dB SPL), with the improvements ranging from approximately 10 to 13 percentage points (Balakrishnan et al, 1996). A characteristic finding in individuals with sensori-neural hearing impairment, in addition to an increase in hearing threshold, is essentially normal loudness discomfort levels, such that there is a reduction in dynamic range. This reduction in dynamic range creates an abnormal relationship between loudness and intensity called loudness recruitment. This reduction in dynamic range is of considerable importance and is applicable to research in C-V ratios, and hearing aid compression design.

Cochlear hearing impairment is typically associated with significant degradation of frequency selectivity (Faulkner and Rosen, 1999). The partial loss of spectral detail appears to have rather slight effects on the perception of connected speech at least in quiet (Baer and Moore, 1993; Shannon et al, 1995). This suggests that temporal cues and changes of gross spectral shape may be of greater significance in speech perception of sentences (Faulkner and Rosen, 1999).

Many cognitive factors determine our ability to understand what has been said. Of course, most essential is our familiarity with the language spoken (Plomp, 2002). Intelligibility depends upon the nature of the speech material used, and a general rule is that performance improves as the speech stimuli become more redundant in any of a variety of ways (Egan, 1948). Miller and colleagues (1951) found that words were more intelligible in a sentence than when presented alone, which also reflects redundancy afforded by the context surrounding the word. Other factors that can contribute to an elderly hearing-impaired person's understanding of speech include acoustical conditions in the environment, memory and cognitive function, talker rate or clarity, and auditory processing deficits. Wingfield et al (2001) considered factors such as cognition, slowing of perception, limitations in working memory, reduced attention, combined with hearing impairment in speech understanding. It has been found that when participants can alter the duration of a sentence that older adults will prefer a longer time than younger adults (Wingfield et al, 2001). Many clinical instruments have been designed to assess speech perception in relation to the above mentioned factors.

Two clinical instruments that will be used in the present study are typical of those used to evaluate speech perception, the Speech in Noise Test (SPIN) and Word Recognition Scores (WRS). The SPIN test (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984; Kalikow, Stevens, & Elliot, 1977) makes use of the concept of redundancy and semantic context by controlling the predictability of the target word, which is always the final word in the SPIN sentence. For half of the sentences, the final word has high predictability, meaning that the recognition of the target word is aided by surrounding contextual cues. The other half are rated as having low predictability, where the listener receives only minimal contextual cues, and therefore must rely on the acoustical properties of the target word for correct recognition. These sentences are presented monotically in the presence of multi-talker babble.

Word recognition tests have been implemented in standard audiometry since the Harvard Psycho-Acoustic Laboratory (PAL) developed 20 lists of "phonetically balanced" (PB) monosyllabic words during World War II (Egan, 1948). The CID W-22 lists (Hirsh et at, 1952) will be used in this study to assess speech discrimination ability for all participants. In addition to the one-syllable characteristic of these test words, the words used in these lists are all fairly common. To make the word-recognition test representative of normal speech, all the sounds are represented about as often as they would normally occur in conversation (Miller, 1951). The CID W-22 lists were developed in 1952 by Hirsh and colleagues and were modeled after the original PB-50.

Consonant versus Vowel Information Research

Caramazza et al (2000) report that there are two distinct views on vowels and consonants. One view is that vowels and consonants are two separate entities that are used to construct syllables for speech production. The other view is that vowels and consonants are just labels used to distinguish intensity differences between peak (vowel) and non-peak (consonant) parts of a continuous stream of sound, Caramazza et al (2000). Caramazza studied the performance in speech production of two Italian-speaking aphasics who had complimentary difficulties producing vowels and consonants. Caramazza et al concluded that vowels and consonants are processed by distinct neural mechanisms, which supports the view of their independent status in language production. The present study is more concerned about the perception of vowels and consonants than the production of vowels and consonants.

The inspiration for the design of this experiment evolved from an article by Cole and colleagues (1996). Cole used the TIMIT database and sorted the phonetic segments into three groups of sounds, consonants, vowels, and weaksons. [The weakson (weak sonorants) class consisted of liquids ([l], [r], [el]), glides ([w], [y]), and nasals ([m], [n],

[nx], [ng], [em], [en], [eng]) (Cole et al, 1996).] Cole's first experiment was designed to (a) remove the spectral information associated with one of the three sound groups by replacing it with either white noise or a periodic sound composed of sinusoids with frequencies ranging from 200 to 4000 Hz; (b) retain the energy and duration profile of the replaced segment; (c) balance the numbers of occurrences of consonants and vowels in each sentence; and (d) balance the total duration of vowel and consonant segments across all sentences (Cole et al, 1996).

Cole et al used sixty sentences from dialect region 2 (DR2) of the TIMIT database. Participants consisted of thirty-five high school students with no reported hearing problems. Participants listened to the speech through headphones at a comfortable volume level set by the subject. Participants could listen to each sentence up to five times at their own pace and entered any English words heard into a computer. The results of the first experiment by Cole et al revealed that almost twice as many words were recognized when vowels (in addition to weaksons) (87.4%) were present opposed to when consonants (in addition to weaksons) (47.9%) were present (Cole et al, 1996).

One possible explanation for the above results was that leaving the weak sonorants in place somehow was more beneficial to the vowels than to the consonants (Cole et al, 1996). A second experiment was conducted by Cole to assess the relative contribution of consonants, vowels, and weak sonorants to word recognition. In this second experiment, listeners could identify 56.5% of the words when vowel information was available, compared to 14.4% for consonants and 3.1% (chance) for weak sonorants (Cole et al, 1996).

Another explanation for the better performance with vowels opposed to consonants in word recognition is that the co-articulatory information available in vowels provides enough information about the surrounding consonants to enable the listeners to recognize the intended words (Cole et al, 1996). To examine this, Cole replicated Experiment 1, but added four new experimental conditions, in which the vowel or consonant segments were either expanded or reduced by moving segment boundaries by 10 msec in each direction before replacing the segments with noise. The performance from expanding consonants resulted in a significant improvement in word recognition whereas expanding vowels improved recognition but not significantly. These results supported Cole's hypothesis that the edges of vowels contain more information about the neighboring consonants than the edges of consonants contain about neighboring vowels. The important result was that even when phonetic boundaries were substantially altered, the direction of the effect (as it affected error rate) was as expected for both vowels and consonants (Cole et al, 1996).

The purpose of the present study was to compare the contribution of information provided by vowels versus consonants to sentence intelligibility when hearing is compromised in the elderly with typical hearing loss. The noise replacement paradigm developed by Cole et al (1996) appealed to us for several reasons: (a) listening involved sentences as test materials; (b) the finding that vowel information improved word recognition scores by a factor of 2:1 compared to consonant information was robust over six experimental conditions. Sentences were presented in three conditions, with either the vowels or the consonants replaced with speech shaped noise, or unaltered. More specifically, the differences in performance in each condition between two groups of listeners, in young normal-hearing (YNH95) and elderly hearing-impaired (EHI95), at a signal level of 95 dB SPL, was examined. In a second experiment, a lower signal level of 70 dB SPL was used with additional young normal hearing listeners in order to compare results from our procedures with those used by Cole et al (1996).

In the present experiment, only two groups of sounds were used, either consonants or vowels. Weak sonorants were classified as a consonant sound in this study. It was hypothesized that: (a) both young normal-hearing listeners and elderly hearing-impaired listeners would perform better with unaltered sentences compared to sentences with either vowels or consonants replaced by noise; (b) young normal-hearing listeners would perform better than the elderly hearing-impaired listeners; (c) young normal-hearing listeners would be able to identify more words correctly when vowels were available compared with consonants available. In addition, performance by (EHI95) listeners were analyzed in relation to clinical measures such as, Word Recognition Scores (WRS), Speech in Noise Test (SPIN), and various pure-tone averages using correlational analyses.

Chapter II. Method

A. Overview of Experimental Design

The contribution of vowel versus consonant information was investigated using a noise replacement paradigm with sentences from the TIMIT (www.ldc.upenn.edu) database. Two listener groups were used in the primary study, one group consisted of young normal-hearing participants (YNH95) and the other group of elderly hearing-impaired participants (EHI95). The signal level was calibrated (Section E) to a high level (referred to as the ‘95 dB’ level) so that the sentences would be reasonably audible for this hearing-impaired group. This presentation level was chosen because it is near the upper limit of comfortable loudness for both the normal-hearing and elderly hearing-impaired listeners. At this presentation level, speech was supra-threshold for both the normal-hearing and the elderly hearing-impaired listeners who met our audiometric criteria (Section B). A second, smaller study was conducted at a conversational level (70 dB SPL) with different young normal-hearing listeners (YNH70) to serve as baseline condition. Overall, this study is a mixed design with groups (YNH95 and EHI95) as a between-subject factor and three phonetic conditions (a) unaltered TIMIT sentences (Full); (b) sentences in which all of the vowels were replaced by noise (Cin); and (c) sentences in which all of the consonants were replaced by noise (Vin), as a within-subject factor. Participants listened for only one testing session which lasted between 1-1/2 to 2 hours in length.

B. Participants

One hearing-impaired (N=16) and two normal-hearing (N=16, N=8) groups, served as listeners for this study. The participants were all American-English native speakers from the North Midland dialect region (DR3) (Garofolo et al, 1993), including Indianapolis and north. A speaker’s dialect region was defined as the geographical area of the U.S. where he or she lived during their childhood years (age 2 to 10) (Garofolo et al, 1993). Participants in the normal-hearing group were all students at Indiana University, ranging in age from 20-35 years of age, and were paid to participate. All participants had pure-tone thresholds of 20 dB HL or better, at octave intervals from 250 Hz through 8000 Hz, and normal tympanometric tracings at the onset of testing.

The elderly hearing-impaired participants were all recruited through the Indiana University Hearing Clinic and were also paid to participate. Their ages ranged from 65 to 80 years (m = 72.9 years). Each subject had a long standing bilaterally symmetrical, moderate sensori neural hearing loss, believed to be of cochlear origin based on acoustic reflex testing, case history, normal tympanograms, and symmetry of air and bone conduction thresholds. Average thresholds for .25, .5, 1, 2, and 4 kHz were 29, 32, 37, 48, and 57 dB HL, respectively. Figure 1 shows individual pure-tone thresholds for the sixteen participants at octave intervals from 250 Hz through 8000 Hz.

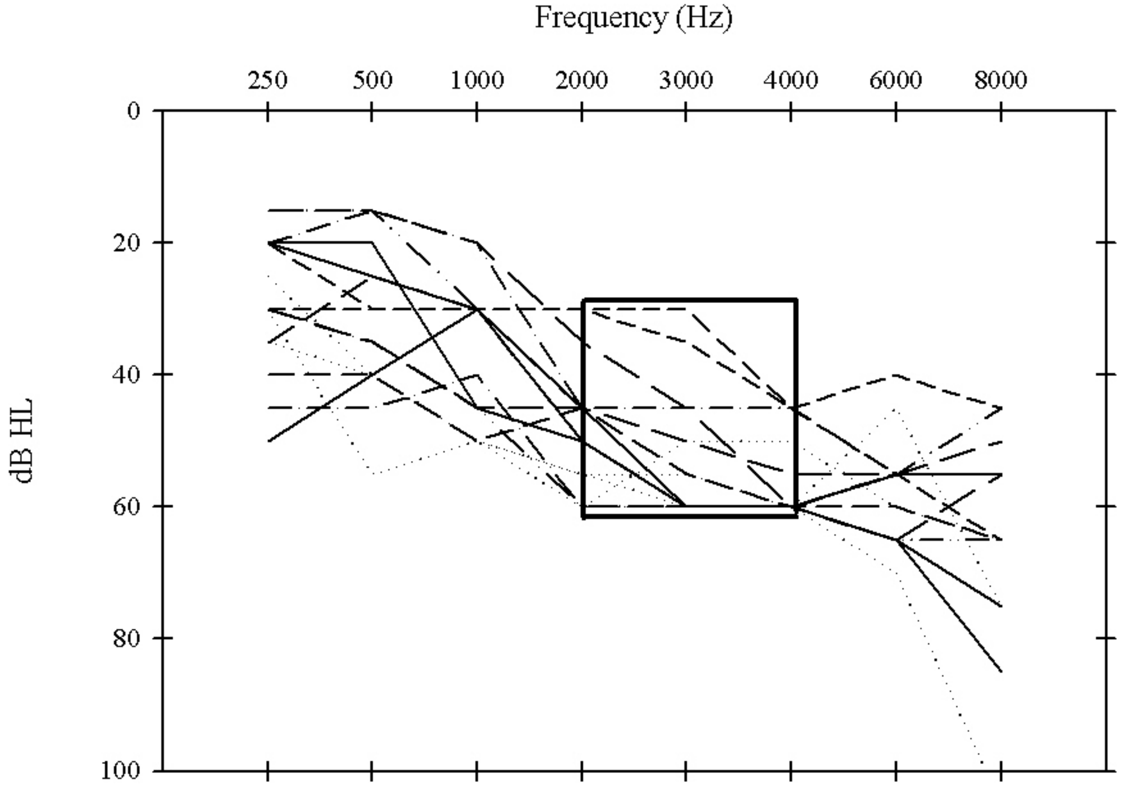

Figure 1. Hearing thresholds for the 16 subjects in the EHI95 group. The box indicates the hearing criteria for this study.

The goal was to choose hearing criterion to include some impairment for vowels at 2000 Hz, while providing audibility for the consonants between 2000 and 4000 Hz. Audibility was estimated using acoustic measurements. The ANSI (1969) standard long-term speech shaped spectrum (LTSS) was selected to represent the average speech spectrum because we wanted to preserve the natural shape of the spectrum for speech. An FIR filter was carefully designed to match the LTSS spectrum (SSS5.b). This filter was applied to a white noise that was created using the ‘randn ()’ function in MatLab to generate a speech shaped noise. Using the same calibration procedures described in Section E below, the 1/3 octave levels of the noise were calibrated to match the low-frequency spectrum of the calibration vowel when it was set to the modal level of vowels in the sentences (95 dB SPL). Given that all signals were low-pass filtered at 4000 Hz (Section E), the criteria for including an EHI95 listener were set at 2000 and 4000 Hz as follows. The 1/3-octave-band level for the noise above normal-hearing threshold at 2000 Hz was 85.6 dB HL and at 4000 Hz was 67.36 dB HL. The minimum loss that might have an effect on vowel perception was estimated to be 35 dB HL. Thus the final criteria for the hearing status of the EHI95 group were thresholds between 35 dB HL and 60 dB HL at 2000 and 4000 Hz, yielding audibility for modal speech of at least 25 dB at 2000 Hz and 7 dB at 4000 Hz. The right ear met the criteria for all participants and was chosen as the test ear. It has been known for some time that speech and language are processed in the left hemisphere of the brain. Because of the anatomical properties of the auditory system right-ear input is represented more strongly in the left hemisphere.

A clinical evaluation was administered to all participants including standard pure-tone audiometry and tympanometry. All participants passed the Mini-Mental Status Examination (MMSE) (Folstein et al, 1975), and a forward and backward digit span test to be included in this study. Because the experiment involved repeating sentences back and identifying words correctly, it was necessary to eliminate memory deficits as a contributing factor. The MMSE and the forward and backward digit span tests were used in order to eliminate any short term memory deficits. The following speech recognition measures were given in the test ear: Speech Reception Thresholds (SRT), Word Recognition Scores (WRS), and 50 SPIN sentences. The SRT is a measure of sensitivity (threshold) for spondaic words (ANSI, 1996). Threshold in this study is defined as the intensity required for 50% recognition of the spondaic words. Word-recognition abilities were assessed in quiet using CID W-22 full lists (50 words) at a presentation level of 40 dB SL [re: Speech Reception Threshold (SRT)]. The SPIN sentences contained 25 high predictability and 25 low predictability sentences that were presented at a level of 95 dB SPL with a signal to babble ratio of +8 dB. These speech-recognition instruments were chosen because of their common usage in various clinical settings.

C. Stimuli

All stimuli were played through TDT system II hardware and ER-3A insert earphones at a sample rate of 16,000 Hz. The stimuli used in this experiment were extracted from the Texas Instruments/Massachusetts Institute of Technology (TIMIT) corpus of read speech (Garfolo et al, 1993, www.ldc.upenn.edu). This database contains speech from 630 speakers representing eight major dialect divisions of American

English. The TIMIT corpus includes time-aligned orthographic, phonetic, and word transcriptions, as well as speech waveform data for each spoken sentence recorded using a 16,000 Hz sample rate. The sentences were categorized in three subgroups: dialect (SA), phonetically-compact (SX), and phonetically-diverse sentences (SI). The SA sentences were meant to expose dialectal variants of the speakers and were read by all 630 speakers. The SX sentences were hand designed to provide a good coverage of pairs of phones, with extra occurrences of phonetic contexts thought to be either difficult or of particular interest. Each of the SA and SX sentences were spoken by several talkers, in order to provide a feeling for speaker variation. Finally, the SI sentences were selected from existing text sources – the Brown Corpus (Kuchera and Francis, 1967) and a collection of dialogs from recent stage plays (Hultzen et al, 1964) so as to add diversity in sentence types and phonetic contexts. Each of the SI sentences were spoken by a different speaker. It was the set of SI sentences that were chosen for this experiment. Sentences were extracted from the TIMIT database from the North Midland dialect region, DR3, which was the best match for listeners from Indiana, Indianapolis and further north. This dialect region consisted of 79 male and 23 female talkers. From sentences containing 6-10 words each, 48 sentences from the TIMIT DR3 corpus were selected randomly with an average of 8.16 words per sentence. The test sentences consisted of 21 sentences each from male and female talkers, and the remaining six sentences were used as training/familiarization sentences. The 42 test sentences were divided randomly into three phonetic conditions, 14 sentences in each condition (see Table 1 and Appendix B). One sentence was picked from each talker, half male, and half female (21 each). Each condition had about 114 words that averaged four phones/word. There were a total of three conditions:

a. “Full” had full phonetic information, i.e. are unaltered sentences.
b. “Vin” had the consonants replaced by noise.
c. “Cin” had the vowels replaced by noise.

	Number of Sentences	**Number of Sentences**	**Number of Words**
	Male	**Female**	
Full	7	7	115
Vin	7	7	114
Cin	7	7	114
Total	21	21	343

Table 1. Distribution of sentences over gender and conditions

D. Sentence processing procedures

Using the TIMIT phonetic transcriptions (Garofolo et al, 1993), their symbols were sorted into 32 consonants and 20 vowels for this experiment, as shown in Table 2.

Group	Phone	Total
Consonant	b d g p t k dx q jh ch s sh z zh f th v dh hh hv l r y w el eng nx m n ng em en	32
Vowel	iy ih eh ey ae aa aw ay ah ao oy ow uh uw ux er ax ix axr ax-h	20

Table 2. Classification of TIMIT phonetic symbols into consonants and vowels

In the altered sentences, either the vowels or the consonants were replaced with a speech shaped noise. The noise shaping was the ANSI (1969) standard idealized long term average speech spectrum, flat from 0-500 Hz, and -9dB/oct roll-off. The FIR filter (SSS5.b) was designed for this shaping in MatLab. The SSS5.b filter was applied to a uniform noise that was generated in MatLab (randn () function). The signal was then measured through the ER-3A insert earphones in a 2-cm^3 coupler with the Larson Davis LD2800 spectrum analyzer, and a –9 dB/oct roll-off was obtained. To set the level of the replacement noise relative to the overall average of the phonemes in the sentences, our goal was to preserve level differences obtained separately for the consonants or vowels. To measure the level differences, eight sentences from the SI group were selected (4 male and 4 female talkers). First the silence was edited out of all the sentences. Then using the phonetic boundaries provided by the TIMIT database, the average level of either the consonants or the vowels was calculated. The difference in the average level of the consonants relative to the vowels (unfiltered, no silence) was calculated as -10 dB. Because the replacement noise is a speech shaped noise, the noise level was referenced to the average level of all the vowels in the sentences in just the 0-500 Hz passband (avl). Next, informal listening was used with various levels of the replacement noise for vowels to determine that the consonants were not severely masked. For example, equating the noise level to (avl) was clearly too loud. Eventually, a level of 6 dB less than (avl) was selected for the vowel replacement noise, which meant that the consonant replacement noise level was 16 dB less. Thus, the duration of consonants and vowels is preserved, and the average replacement noise level is at least 6 dB less than the average levels measured for the vowels or consonants in these sentences.

The boundary locations used for noise replacement were those provided in the TIMIT corpus in the .phn files with the exception of our following three rules, which were used in less than 5% of the altered sentences:

1. Stop closure symbols (i.e. “bcl”) were combined with stop (i.e. “b”) and treated as a single consonant. This yielded the 32 unique consonants in Table 2.
2. Syllable final V + [r] as in “beer” were transcribed in TIMIT as a vowel plus the consonant [r]. These V+[r] transcriptions were considered as a single rhotocized vowel in this study.
3. The glottal stop [q] was a separate consonant in TIMIT, generally realized by silence. When the glottal stop was transcribed as occurring between two vowels, [VqV], it was treated as a vowel in this study.

E. Calibration

The first principle of the calibration method was to ensure that no speech segments would be presented at a level greater than 100 dB SPL. First, all sentences were normalized to the average RMS value across all 42 sentences. The loudest vowel

(LdV_i) in each sentence, i, was measured digitally as the RMS value for a 50 ms window located around the peak amplitude. There were two vowels at this level and both of these vowels were highly stressed within their sentences. The vowel /uh/ was picked from the word ‘study’ as the calibration vowel (calvow4.wav) to adjust the system gain to the 100 dB SPL level. The RMS levels for the other LdV_i values ranged over 11 dB. However, the distribution of the LdV_i showed a narrow peak (mode) such that 71% of the LdV_i values were within +/-2 dB of the peak.

A low-pass FIR filter was designed to avoid variability in high-frequency audiograms for the elderly hearing-impaired with the sloping hearing losses. The filter was flat until 4000 Hz, had the -3 dB cutoff at 4400 Hz, and then a steep slope that fell 50 dB by 5000 Hz (>200 dB/octave). Using the filtered calvow4.wav vowel, level was calibrated through ER-3A insert earphones in a 2-cc coupler using the Larson-Davis LD2800 spectrum analyzer and the linear weighting scale to be 100 dB SPL. Given that the maximum level for the TDT equipment was 113 dB SPL, the 100 dB level did not have any distortions. The sentence with the calvow4.wav (from the word ‘study’) was not too loud in informal listening. With this calibration, the level for the modal peak of the LdV_i was 95 dB SPL.

Background noise was used in this study to mask any sentence editing transients. This uniform noise was generated by the TDT WG2 and was also filtered by the 4400 Hz low-pass filter. It was calibrated using the earphones and equipment above to have a level less than -50 dB re the 100 dB SPL calibration level.

F. Procedures

Stimuli were presented at the 100 dB SPL calibration level (most vowels near the modal 95 dB SPL level) to the right ear of all participants. Each sentence was presented twice at the participants own pace. Participants were asked to verbally repeat as many of the words that could be identified. Before testing began, simple instructions were given:

> *You will be listening to 42 different sentences, some of the sentences will have either vowels or consonants replaced with some noise, so they may sound a little "choppy". Each sentence will be played twice for a total of 84 presentations. Your job, after each listen, is to repeat all of the words you think you heard. I will be scoring the words that you get correct. So, if you don't understand the meaning of the sentence, that's fine. But, I want you to do your best to repeat all the words that you hear. If you are not sure, I encourage you to guess*

Listeners practiced with six sentences that were not in the test set, with feedback before beginning the experiment. This was done so that participants had an idea of what to expect before starting the experiment, and to reduce any learning effects that could occur using these noise altered sentences. Two randomizations of the sentence list order were used to reduce order effects with half of the participants in each group listening to each test order. Prior to testing, all of the sentences were typed out and the experimenter scored the responses by hand by circling words that were identified correctly. A recorded backup from a digital recorder was also made to assist in scoring.

Chapter III. Results

General results

The primary goal of this experiment was to determine how well normal-hearing and hearing-impaired listeners identify words in sentences, both unaltered and with only consonants or only vowels, when the sentences were audible. To compute a score for each sentence, the words had to be identified in their entirety. The semantic meaning of the sentences was not considered in this experiment, only correct word recognition. Participants were encouraged to guess at what they thought they heard if they were not completely sure what was said. The total number of words correct for each listener and sentence condition (Table 1, Methods) was used in the statistical analyses. The total number of words identified correctly in each condition was divided by the total number or words in each condition and converted to percentage values for tables and figures.

	Full mean	Vin mean	Cin mean
YNH	99	65.1	51.6
EHI	93.8	40.2	20
Combined	96.4	52.65	35.8

Table 3. Percentage of words identified correctly in each condition.

To provide an overview of the performance for the different sentence conditions, scores were averaged over listener groups and are presented as percentage of words correctly identified in Table 3 and Figure 2. Overall, as expected, both groups together performed significantly better in the Full condition compared in Vin and Cin. Clearly, both the normal-hearing and the elderly hearing-impaired groups demonstrated no difficulty identifying words in the unaltered sentences at 95 dB SPL. Ceiling performance (100% correct) was seen for 8 of the 16 YNH95 participants. Given that the scores for the Full condition were high for both groups, we examined whether they were significantly different using a two-way analysis of variance (repeated measures) with two listener groups and three conditions as the repeated measures factor and number of words correct as the dependent variable [$F(2,60)=23.844$, $p<0.001$; see Fig. 2]. For the Full condition, YNH95 and EHI95 listeners performed the same according to the Tukey HSD post-hoc analysis ($p<0.91$). Similar results were obtained in a separate ANOVA when the number of words correct was transformed to rationalized arcsine units (RAU) to stabilize the error variance (Studebaker, 1985). It should be noted that only one of the participants in the EHI95 group scored poorly, 67.8% correct, on the ‘Full’ condition whereas the range for the other EHI95 listeners was 88.6% to 99.1 % correct such that there was considerable overlap with the YNH95 group (range 97.4% to 100%). The fact that these two groups scored equally well for the ‘Full’ condition verified that the sentences were fully audible for the EHI95 group and the listening/repeating task was not difficult to perform. The focus of this experiment was to examine the contributions of vowels and consonants to sentence understanding. Because both groups performed very well (m=96.5%) in the ‘Full’ condition, it was not considered in further analyses.

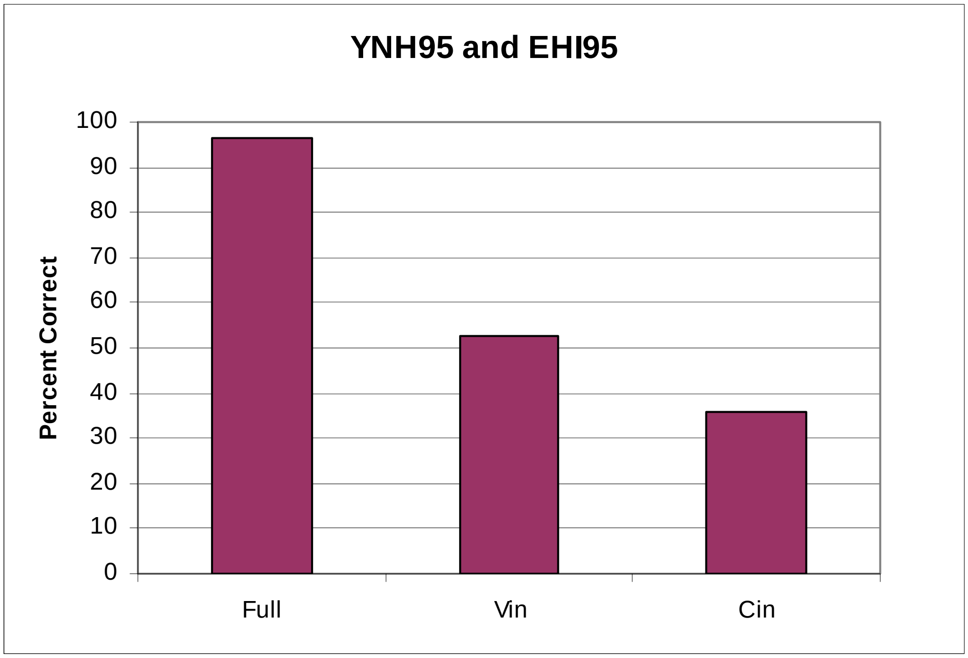

Figure 2. Total number of words correct in each condition for both YNH95 and EHI95 in percent correct

Comparison of YNH95 and EHI95 Listeners

Performance for the two listener groups to identify words in the vowel only (Vin) and consonant only (Cin) conditions is shown in Fig. 3. As expected, the YNH95 group (m = 58.4%) performed significantly better overall than the EHI95 group (m = 30.0%) according to a two-way analysis of variance with two listener groups and two conditions as the repeated measures factor (F(1,1)=52.9, p<0.001). The performance between groups yields a 28.4% difference in overall scores between YNH95 and EHI95. This represents an overall decrease in performance for the EHI95 group for both Vin and Cin conditions because there was no significant interaction [F(1,30)=2.45, p=0.128; see Fig.

3]. Interestingly, both groups performed significantly better with the Vin condition (m = 52.6%) than the Cin condition (m = 35.6%) [F(1,30)=62.1, p<0.001; see Fig. 2].

In the YNH95 group, percent-correct performance in the Vin condition (65.1%) was significantly higher than the words correct in the Cin condition (51.6%) according to the Tukey HSD post-hoc analysis (p<0.001). Also, in the EHI95 group, percent-correct performance in the Vin condition (40.1%) was significantly higher than in the Cin condition (19.9%) according to the Tukey HSD post-hoc analysis (p<0.001).

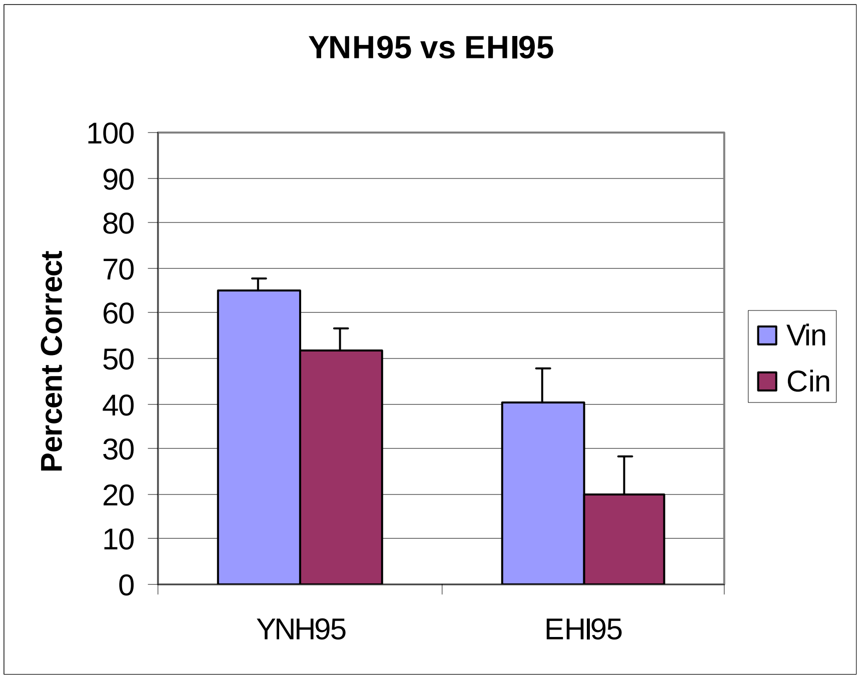

Figure 3. Number of words correct by YNH95 and EHI95 in Vin and Cin conditions. Error bars indicate standard error.

Comparison of Signal Level for Young Normal-Hearing Listeners

To examine any effects of the high signal level on the YNH95 group, a second experiment was conducted at a level that is slightly louder than normal conversational loudness (70 dB SPL). Cole et al.'s (1996) data were obtained at the participants' most comfortable level (MCL) which is likely substantially lower than our presentation level in the primary experiment (95 dB SPL). A different group of young normal-hearing listeners (YNH70) (N=8) was run at a lower level of 70 dB SPL. Comparisons were made between the YNH70 group and the first eight participants in the YNH95 group (Fig. 4). According to a two-way analysis of variance with two listener groups and two conditions as the repeated measures factor, there was no overall difference in performance (m=58.4% at 95 dB, m=54% at 70 dB) between presentation levels, however there was a significant interaction [$F(1,14)=3.15$, $p<0.097$]. According to the Tukey HSD post-hoc analysis, the YNH95 group performed significantly better (51.6%) in the Cin condition compared with the YNH70 group (34%) ($p<0.005$). Perhaps this is due to the increased audibility for the consonants in the louder presentation level. There was no significant difference in performance (See Fig. 4) between groups in the Vin condition ($p=0.38$). Thus, the difference between the low and the high signal levels can be described as only affecting the Cin performance, which significantly improved by 15.6% at the higher signal level. Because the higher signal level increased performance in the Cin condition significantly, the effect of masking by the replacement noise for vowels on the consonants was not increased for young normal hearing listeners.

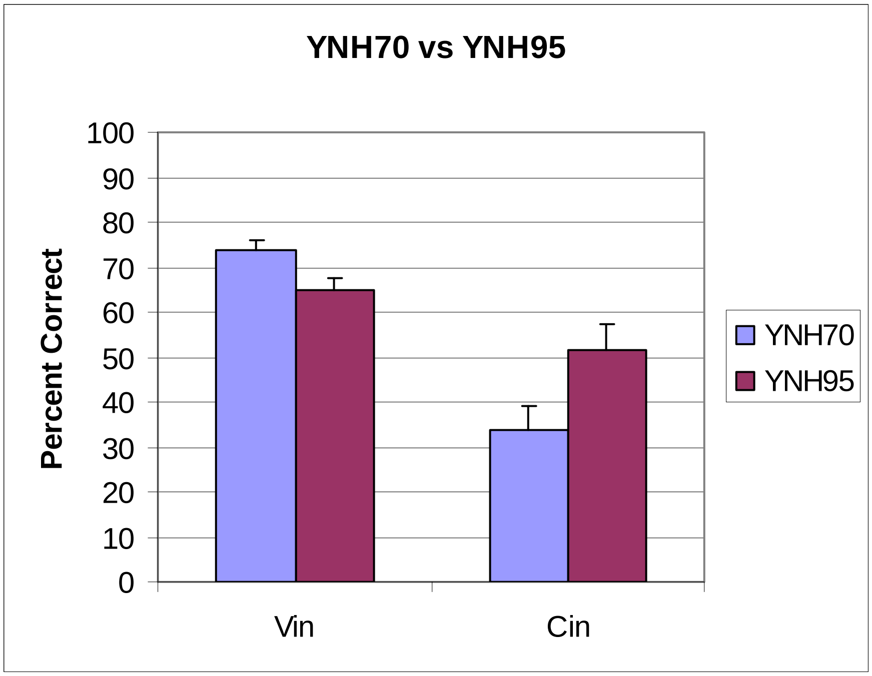

Figure 4. Number of words correct by YNH70 and YNH95 in Vin and Cin conditions. Error bars indicate standard error.

Chapter IV. Discussion

General

As expected, all three groups of listeners could identify more words correctly when sentences were presented without any alteration. Because all three groups performed accurately with the unaltered sentences, it is believed that the TIMIT sentences were an adequate choice for the stimuli, and the task was not inherently difficult to perform. Within the noise conditions (Vin, Cin), as expected, the YNH95 performed substantially better than the EHI95 group. It was hypothesized that the YNH95 group would perform better when vowels were available as opposed to when consonants were available based on Cole et al.'s (1996) experiment. However, it was not easy to predict how well hearing impaired listeners would perform. The 2:1 advantage of Vin over Cin conditions for the EHI95 listeners makes it clear that the information carried by the vowels contributed more to intelligibility in sentences than did the consonants for both groups of listeners.

Within the EHI95 group, although there was a wide distribution of scores for the Cin condition (0% - 54.4%), the mean for the Cin condition was 19.9% which is still well above chance given that word identification in sentences is an open response task. However, four of the EHI95 participants did score less than 5% correct for the Cin condition which is estimated to be chance. Without these four participants the mean for the Cin condition was 25.8% correct, the vowel benefit even more similar to the YNH95 group. Thus, the consonants were not completely masked for the majority of the EHI95 participants, per our design. It is possible that the level of cognitive function, educational level, and level of interest in this study influenced the degradation in performance in both conditions (Vin, and Cin) despite the passing scores on the Mini-Mental Status Examination as well as the digit span test.

Figure 3 clearly shows that there was no interaction between YNH95 and EHI95 and the noise replacement conditions. Thus, the YNH95 participants also found the Cin task more difficult, even though they performed the Cin task better at the higher, 95 dB level compared to 70 dB SPL. (Note that the EHI listeners were not run at the 70 dB SPL level because most of the consonants would have been inaudible. Thus it is assumed that there would also be improvement by the EHI95 group from near chance performance at 70 dB to the overall score of 35.7% obtained at 95 dB SPL). However, the result that both groups had nearly the same decrease from Vin to Cin (no interaction) implies that there was no extra masking or distortion for the EHI95 participants due to their hearing impairment.

It was surprising to find a significant improvement by the YNH listeners in the Cin condition with the higher presentation level in the second experiment. Perhaps this improvement at the 95 dB SPL level is due to the increased sensation level of the consonants when presented at a louder signal level. Note that consonants should be fully audible at 70 dB SPL (+45 dB SNR re: the background noise), so the nature of the changes in consonant information or processing at the higher level is not known. Interestingly, comparing both YNH groups, the presentation level did not significantly affect the performance within the Vin condition. If anything, the 9% decrease in vowel performance at the higher level might indicate some distortion of the vowels in the auditory system.

Comparison to Previous Research

The purpose of the second experiment was to replicate some of the results reported by Cole et al. (1996). Although the two experiments differed in many details, such as choice of sentences, choice of noise and noise level, and assignment of TIMIT phonetic symbols to consonant or vowel classes, the similarities of the results between the two experiments are striking. In several different noise replacement experiments by Cole, higher scores in the Vin condition along with a large difference in performance between the Vin and the Cin condition were obtained. Those results were replicated for all three groups, YNH95, EHI95 and YNH70. The benefit ratio between Vin and Cin in Cole’s experiments was usually a factor of 2:1. This was the same factor observed in two

of our three of our groups; namely the YNH70 (ratio 2.2:1) and the EHI95 (ration 2.0:1) groups. For the YNH95 group, however, the ratio was substantially lower, about 1.26:1. The YNH70 group was run specifically to replicate one of Cole et al.'s experimental conditions; namely, his Experiment 1 when vowels or consonants were replaced by white noise. Our results showed that the Vin condition was significantly higher (74%) than the Cin condition (34%) ($p<0.001$). The difference between Vin and Cin was 40% in our experiment, which was very close to Cole's Experiment 1, where the difference between Vin and Cin was 40.8%. Interestingly, Cole's participants performed 13% better in both the Vin and Cin test conditions that were most similar to ours. This could be due to presentation level differences. As stated earlier, Cole's participants had control over their own presentation level. It is difficult to determine the presentation levels listeners selected, but they may be assumed to be close to our 70 dB SPL (loud conversational speech). Note that in our experiments the vowel benefit decreases with higher presentation levels, so this is an uncontrolled variable in Cole et al.'s experiment.
The present study, as well as Cole et al. (1996), focused on the contributions of vowels and consonants to fluent speech (i.e. sentence intelligibility).

Our tasks involved listening and repeating sentences based on either vowel information or consonant information alone. In contrast to our experiments, the experiments by Fletcher (1929) and Miller (1951) concentrated more on the "articulation" of vowels and consonants in single monosyllabic words, or non sense syllables over a transmission system. These experiments were devised to develop a better transmission system to communicate over long distances. Fletcher found that more errors were made with the consonants, and thus it was believed that consonants contribute more to speech intelligibility. However, we note two primary differences between the early and present experiments. First, linguistic processing of monosyllables relies on bottom-up information, while our sentence intelligibility task incorporates considerable predictive information from top-down processing. Second, the interpretation of the outcomes of the two types of experiments differs. The phoneme error analysis in Fletcher and in Miller examined discrete symbols from only the monosyllables incorrectly identified and cannot reveal what the role of acoustic information distributed across syllables (coarticulation) had in the correctly identified monosyllables or phonemes. The noise replacement task removes specific acoustic information and directly measures the accuracy of word identification from the remaining distributed information. Thus our conclusion that vowels carrying more information than consonants relates the presence of acoustic information in signals as processed in the auditory system without an intervening symbol analysis.

Let's consider whether this implies that the location of the acoustic boundaries in the noise replacement task is critical to our conclusion. In their third experiment, Cole et al. (1996) altered segment boundaries by ± 10ms to examine the contribution of coarticulated information about consonants and vowels. As expected, longer consonantal signals yielded higher word identification in the Cin condition than shorter ones. However, word identification for the correspondingly shorter vowel signals in the Vin condition was still greater (79.8%) than for longer consonants (53.1%).

Thus, overall, our conclusions are in agreement with the general results of Cole et al. (1996) and somewhat at odds with the interpretations of the earlier experiments by Fletcher (1929) and Miller (1951). The acoustic information distributed across the vowels in speech carries more information important for accurate sentence processing than do consonants, regardless of the details of the segment boundaries selected to separate consonants from vowels. This is true for both normal-hearing and hearing-impaired listeners.

Relation to Clinical Tasks

Clinical testing was done to assess the participant's ability to understand speech in both quiet and in noise with standard clinical instruments and to relate these scores to the scores obtained in our noise replacement experiment. Each subject in the YNH95 and the EHI95 groups was tested with two standard clinical audiologic measures, the Speech in Noise Test (SPIN) and Word Recognition Scores (WRS). As expected, for each measure, the YNH95 performed better than the EHI95 group. The SPIN test has three scores, high probability, low probability, and the combination score of both. The YNH95 group performed better on all scores: 96.5% versus 90.8% on the high probability SPIN sentences; 56% versus 34.6% on the low probability sentences and 76.3% versus 62.7% on the combined scores. Word recognition was assessed using CID W-22 full lists at a presentation level of 40 dB SL (re: Speech Reception Thresholds). The YNH95 group again performed better (98.1%) than the EHI95 group (88.9%) on WRS.

For the elderly hearing-impaired listeners we examined the relationship between our novel noise replacement tasks and standard audiologic clinical measures, including audiometric thresholds using correlation and regression. Because there were many measures near 100% (SPIN High probability sentences and the word scores for Full sentences), all percentage scores were transformed to rationalized arcsine units (RAU) to stabilize the error variance (Studebaker, 1985). First product-moment correlation coefficients were calculated between the Full, Cin and Vin RAU scores and the clinical scores for the WRS, SRT, and SPIN High and Low probability sentences as shown in Table 4.

	WRS	*SRT*	*Spin High*	*Spin Low*	*Cin*	*Vin*	*Full*
SRT	-0.10						
Spin High	0.55*	-0.24					
Spin Low	0.42	-0.52*	0.66*				
Cin	0.64*	-0.58*	0.38	0.61*			
Vin	0.74*	-0.33	0.55*	0.66*	0.58*		
Full	0.62*	-0.30	0.51*	0.72*	0.60*	0.64*	

Table 4. Correlations using RAUs between standard clinical instruments and experimental conditions Cin, Vin, and Full (*p<0.05)

All speech perception measures were positively correlated with themselves and negatively correlated with the SRT. The correlation coefficient, r, for an N of 16 must be greater than 0.49 to be significant at the $p<0.05$ level. All three experimental conditions (Full, Vin, and Cin) were moderately and significantly correlated with the WRS. Also, all three experimental scores were moderately and significantly correlated with the SPIN ‘low probability’. The Full and Vin conditions were moderately and significantly correlated with the SPIN ‘high probability’. The SRT was negatively correlated with all three experimental conditions, but only moderately and significantly correlated with the Cin condition.

Correlational analysis was also used to examine the relation between the hearing thresholds in dB HL with performance on all three experimental conditions (see Table

5). For thresholds, most correlation coefficients were negative because a high threshold in dB HL means poorer audibility of the speech sounds. Thresholds at 2000 and 3000 Hz were highly correlated ($r < -0.85$) with the Cin condition and moderately correlated ($r < -0.60$) with the Vin condition. Thresholds at 1000 and 4000 Hz were moderately and significantly correlated ($r < -0.61$) with the Cin condition.

	Cin	***Vin***	***Full***
250	-0.22	-0.20	-0.29
500	-0.44	-0.32	-0.27
1000	-0.62*	-0.20	-0.23
2000	-0.89*	-0.66*	-0.69*
3000	-0.86*	-0.61*	-0.53*
4000	-0.67*	-0.46	-0.39
6000	-0.36	0.03	-0.04
8000	-0.33	-0.25	-0.08

Table 5. Correlations using RAUs between pure-tone hearing thresholds and experimental conditions Cin, Vin, and Full (*$p<0.05$)

To characterize the EHI group, various pure-tone averages were calculated. These values and the subject's age were correlated with the scores for the three experimental conditions as shown in Table 6. The pure-tone average of 2000, 3000, and 4000 Hz and 3000 Hz and 4000 Hz were highly correlated with the Cin condition. An overall pure-tone average of 250, 500, 1k, 2k, 3k, and 4k was also highly correlated with the Cin condition ($r = -0.79$), and moderately and significantly correlated with the Vin condition. The other pure-tone averages that were moderately and significantly correlated with the Vin condition were 2000, 3000, and 4000 Hz, and 3000 and 4000 Hz (see Table 6). There were no significant correlations between participants age and the three task conditions (r<|0.36|).

Pure-tone Average	Cin	Vin	Full
500, 1k, 2k	-0.75*	-0.46	-0.42
3k, 4k	-0.81*	-0.57*	-0.43
2k, 3k, 4k	-0.9*	-0.65*	-0.53*
250-4k	-0.79*	-0.54*	-0.5*
EHI Age	0.09	-0.18	-0.35

Table 6. Correlations using RAUs between pure-tone averages, age, and experimental conditions Cin, Vin, and Full (*p<0.05)

With high correlations observed between the many variables, linear regression analysis was undertaken to determine how the variables were associated with the scores in our three experimental tasks. Forward, step-wise linear regression in Statistica (www.statsoft.com) was selected as the anaylsis tool. Examining the variables in the correlation matrices above we determined which independent variables to include in the linear regression. Seven clinical tests were selected: the measures for the WRS, SRT, SPIN High and SPIN Low sentences; the elderly listeners' ages; and the two pure-tone averages that appeared to represent the speech information, the "PTA Vowel" average of thresholds for 500, 1000, and 2000 Hz, and the "PTA Consonant" average of thresholds 3000 Hz and 4000 Hz.

Consider the linear regression analysis for the elderly listeners' performance on the Full sentences. Although their performance was generally quite good (m = 93%), there was sufficient variance about the mean to conduct an analysis. First, only the Vin and Cin scores were included in the regression to determine if either one contributed significantly to normal sentence recognition. Neither the Vin or the Cin scores contributed significant variance towards predicting the Full score ($p > .09$). Apparently, auditory processing necessary to do the unusual Cin and Vin noise replacement tasks is not similar in kind to processing fluent speech (in the Full task). Therefore, no experimental scores (Full, Cin, Vin) were included in subsequent analyses that therefore focused on only the clinical measures.

In the results of the linear regression between the Full RAU score and the seven clinical variables, only one variable contributed significantly in the regression [$r = 0.72$, $F(1,14) = 14.7$, $p < .002$], the SPIN Low sentence score. With an $R^2 = 0.51$, the variability of individual performance on the Full sentence task that relates to that of the SPIN Low sentence task is 51%. Additional variance was accounted for by WRS (+12%), but it was not significant in the regression. These results are interpreted to mean that for elderly hearing-impaired persons, the identification of words in ordinary sentences, such as the Full sentences, is related moderately to the clinical tasks of scoring one word in a low predictability sentence (SPIN Low), and the identification of single words. Presumably a major source of the variance not accounted for relates to differences in cognitive processing needed to perform these more simple clinical tasks compared to the identification of meaning in fluent speech, as represented by our Full sentence task.

Linear regression was also conducted on the Vin RAU scores for the sentences in which all the consonants were replaced with noise. Of the seven clinical variables, only two variables significantly contributed to the regression, WRS with $R^2 = 0.55$, and the SPIN Low scores with an additional variance of $R^2 = 0.15$ ($F(2,13) = 15.3$, $p<0.001$).

Notably the WRS and SPIN low scores are the same two variables that were associated with the Full sentence task, although the order of their contributions is reversed. This can be interpreted in a similar way, namely that processing of the vowel only information by elderly hearing impaired listeners is more related to abilities to processing sound in other speech tasks, such as identification of individual words, whether in lists (WRS) or SPIN Low sentences, than to audiometric and other variables examined.

The linear regression of the Cin scores presented very different results. The variable associated most with Cin (r = -0.81) was the PTA for the 3 and 4 kHz frequencies, termed the PTA for consonant spectral information. Thus, the poor audibility of the consonantal spectral information is the largest variable associated with understanding sentences with only consonants present. Significant additional variance of +15% was obtained with the PTA of 500, 1000, and 2000 Hz. Another variable identified as contributing to the variance was the SRT (+6% variance, p = 0.043). Audibility related factors were closely associated with performance on the consonant only information task, with the SRT that measures audibility of speech in dB, also a factor. Together the performance by the elderly hearing impaired could be predicted from these factors with 87% of the variance, the highest of the regressions for the three experimental variables.

Another linear regression focused just on audibility was conducted on the Cin and Vin scores. Only the individual hearing thresholds at 250, 500, 1000, 2000, 3000 and 4000 Hz were used as the variables in these regressions. Interestingly, the threshold at 2000 Hz accounts for 80% of the variance for the Cin condition, with an additional 13% contributed by thresholds at 3000, 250 and 500 Hz ($R^2 = 0.93$, $F(4,11) = 32.2$, $p<0.001$), corroborating our findings of the Cin regressions including the pure-tone averages. For the Vin condition, only the thresholds at 1000 and 2000 Hz contributed significantly and accounted for only 67% of the variance ($F(5,10) = 4.17$, $p<0.026$), which approaches the contributions of the WRS and SPIN low of 70%.

A final regression was conducted on the Cin and Vin conditions using WRS, SRT, SPIN high and low, Age, and the high frequency pure-tone average of 1000, 2000, and 4000 Hz. The high frequency PTA of 1, 2, and 4 kHz was used in this regression because of its relevance in the literature. These regressions showed similar results to the regressions with the other pure-tone averages. Namely, for the Vin condition, only two variables significantly contributed to the regression, WRS with $R^2 = 0.55$, and the SPIN Low scores with an additional variance of $R^2 = 0.15$ ($F(2,13) = 15.3$, $p<0.001$). And for the consonants, the only variable associated with Cin was the PTA for the 1, 2, and 4 kHz frequencies ($R^2 = 0.78$).

The results of these regressions suggest that for elderly hearing-impaired, audibility in the higher frequencies can better predict sentence intelligibility from only consonant information compared to the two clinical instruments using speech (WRS, SPIN) that were used in this experiment. Of interest is that variability in the pure-tone thresholds of 2 kHz, 3 kHz, 250 Hz and 500 Hz alone predicted the Cin scores with very high accuracy, accounting for 93% of the variance. Conversely, the WRS and the SPIN-low variables were better predictors of intelligibility from vowels only compared to measures of audibility. Because the vowel only scores were better predicted using speech evaluations, while the consonant only scores were better predicted by audibility, it can be speculated that vowels and consonants are cognitively processed differently by the auditory system. This is a similar finding that was reported by Caramazza et al. (2000) suggesting that vowels and consonants are likely processed by two independent neural mechanisms.

Chapter V. General Summary and Conclusions

The purpose of this study was to compare the contribution of information provided by vowels versus consonants to sentence intelligibility in young normal-hearing listeners and when hearing is compromised in the elderly with a typical hearing loss. The stimulus selection and calibration in this study was done to ensure audibility of some consonants at the high frequencies up to 4000 Hz. It would not be surprising that hearing-impaired participants would perform poorly with the consonant condition if the consonants were inaudible. Presentation at a high signal level therefore enables sensible comparisons between vowels and consonants within this group. Contrary to popular belief by scientists and clinicians alike, both groups performed significantly better when the vowel information was preserved as opposed to the consonant information. This was true even when ensuring audibility for consonantal speech sounds and with durational information preserved in the noise replacement procedures.

The majority of the consonant/vowel research that has been done in the past focused on CV (consonant-vowel), VC (vowel-consonant), and CVC (consonant-vowel-consonant) monosyllabic words and non sense syllables. Very few studies have examined their contributions to fluent speech, such as sentences. Fletcher's early work (1929) in this area was motivated by the need for a transmission system following the first World War. In his experiments, he concentrated on articulations and error patterns made by consonants and vowels. However, we communicate using fluent speech (sentences) and the segmentation, envelope, and durational cues in sentences also aid in speech intelligibility. Using monosyllabic words, or non sense syllables, only assesses a portion of the intelligibility in a spoken message. For example, the correlation between WRS and our Full sentence condition showed that the WRS predicted about 38% of the variance in the Full sentence task.

At high signal levels (95 dB SPL), both groups of participants performed equally well with the unaltered sentences. The fact that both groups were near ceiling (m = 96.4%) implies that the task at hand was not at all difficult to perform, and that the sentences were sufficiently audible for the hearing-impaired listeners to identify words and meaning in the sentences. It also implies that the sentences chosen from the TIMIT corpus were a good choice as stimuli that were not cognitively challenging for the elderly listeners. It was reasonably hypothesized that both groups would perform better with the unaltered sentences compared to sentences with either vowels or consonants replaced by noise. This was indeed the case across groups and also within the young normal-hearing subjects at both signal levels. It is interesting that the EHI95 group, with the exception of one individual (who also scored the poorest on the WRS, the SPIN ‘combination’, and the Vin condition), scored so remarkably well with the unaltered sentences despite a moderately sloping cochlear hearing impairment. This implies that, even with a cochlear hearing impairment, if speech is presented loud enough it can be understood very well. Some would argue that presenting speech at a level of 95 dB SPL would cause some degree of upward spread of masking that inherently would degrade the speech intelligibility. However, in the present experiment, this was not the case.

It was also hypothesized that the young normal-hearing participants would perform better than the elderly hearing-impaired participants. Across all conditions, this was true. As mentioned above, both groups performed equally well statistically in the unaltered sentences, even though the YNH95 averaged 5 percentage points higher. However, in the Vin and Cin conditions, the young normal-hearing participants performed significantly better than the elderly hearing-impaired. If audibility was ensured, the task was not too difficult, and the elderly hearing-impaired listeners performed near ceiling with unaltered sentences, why did they not perform equally with the young normal-hearing participants in the Vin and Cin conditions?

One possible reason for the poorer performance in the elderly is the combined deficits of temporal resolution, frequency selectivity, and loudness recruitment that are associated with moderate hearing impairment. It is possible that the elderly did well with the unaltered sentences because they relied on overall amplitude envelopes, durational cues, semantic context, and appropriate formant transitions to identify words correctly. Most of these cues were severely distorted in the Vin and Cin condition, although the durational cues were largely intact. Therefore because the elderly relied on a degraded auditory system to process these distorted cues, a 20% decrease in performance was obtained compared to young normal hearing listeners.

The most remarkable outcome of this study was the advantage of vowel information to sentence intelligibility in all groups and presentation levels. It was hypothesized that the young normal-hearing participants would indeed do better with vowel information, but it was not known how the elderly hearing-impaired listeners would perform with vowels or consonants alone. Even though the elderly performed poorer as a group, the vowel information contributed significantly more to sentence intelligibility than did the consonant information. Of particular note is that the 2:1 benefit ratio of vowels over consonants reported by Cole et al. (1996) was replicated for the YNH70 group and was observed for the EHI95 group as well.

Overall, the results of this study suggest that vowel information contributes significantly more to sentence intelligibility for both young normal-hearing and elderly hearing-impaired listeners. Although consonants do contribute to sentence intelligibility, clearly across groups and signal levels, vowels are more helpful in understanding fluent speech.

Appendix A. Phonemic and phonetic symbols used TIMIT

SYMBOL	EXAMPLE WORD	POSSIBLE PHONETIC TRANSCRIPTION
------	------------	-------------
Stops:		
b	bee	BCL B iy
d	day	DCL D ey
g	gay	GCL G ey
p	pea	PCL P iy
t	tea	TCL T iy
k	key	KCL K iy
dx	muddy, dirty	m ah DX iy, dcl d er DX iy
q	bat	bcl b ae Q
Affricates:		
jh	joke	DCL JH ow kcl k
ch	choke	TCL CH ow kcl k
Fricatives:		
s	sea	S iy
sh	she	SH iy
z	zone	Z ow n
zh	azure	ae ZH er
f	fin	F ih n
th	thin	TH ih n
v	van	V ae n
dh	then	DH e n
Nasals:		
m	mom	M aa M
n	noon	N uw N
ng	sing	s ih NG
em	bottom	b aa tcl t EM
en	button	b ah q EN
eng	washington	w aa sh ENG tcl t ax n
nx	winner	w ih NX axr
Semivowels and Glides:		
l	lay	L ey
r	ray	R ey
w	way	W ey
y	yacht	Y aa tcl t
hh	hay	HH ey
hv	ahead	ax HV eh dcl d
el	bottle	bcl b aa tcl t EL

Vowels:

iy	beet	bcl b IY tcl t
ih	bit	bcl b IH tcl t
eh	bet	bcl b EH tcl t
ey	bait	bcl b EY tcl t
ae	bat	bcl b AE tcl t
aa	bott	bcl b AA tcl t
aw	bout	bcl b AW tcl t
ay	bite	bcl b AY tcl t
ah	but	bcl b AH tcl t
ao	bought	bcl b AO tcl t
oy	boy	bcl b OY
ow	boat	bcl b OW tcl t
uh	book	bcl b UH kcl k
uw	boot	bcl b UW tcl t
ux	toot	tcl t UX tcl t
er	bird	bcl b ER dcl d
ax	about	AX bcl b aw tcl t
ix	debit	dcl d eh bcl b IX tcl t
axr	butter	bcl b ah dx AXR
ax-h	suspect	s AX-H s pcl p eh kcl k tcl t

Others:

SYMBOL ------	DESCRIPTION -----------
pau	pause
epi	epenthetic silence
h#	begin/end marker (non-speech events)
1	primary stress marker
2	secondary stress marker

Appendix B. Sentences used in each condition

Full Condition:

1. But it did print good verse and good fiction.
2. Crooked, overlapping, twisted, or widely spaced teeth.
3. Originally, the main types used were various compositions of polyesters.
4. Perhaps it was right; perhaps it was just.
5. Two other cases also were under advisement.
6. Trespassing is forbidden and subject to penalty.
7. To honor him is to honor ourselves.
8. It was exposed to a high velocity gas jet.
9. What did you mean by that rattlesnake gag?
10. In a way, he couldn't blame her.
11. Or maybe you just don't feel like a cigar?
12. A warm breeze played across it, moving it like waves.
13. But problems cling to pools, as any pool owner knows.
14. They took it away, overalls or something.

Vin Condition:

1. It was muscular but it wasn't symmetrical.
2. Necessary retouching was put on at once.
3. It helps those people who help themselves.
4. Can thermonuclear war be set off by accident?
5. However, this inaugural feast did its sponsors no good whatever.
6. We always thought we would die with our boots on.

7. There was a thick, squashy crack of fist on flesh.
8. It all did look very efficient and shipshape.
9. In the pity for them his loneliness was gone.
10. View takes in five counties, two bedrooms.
11. If she moved gracefully, she was clumsy at it.
12. What elements of our behavior are decisive?
13. We must be ready for any needed sacrifice.
14. Her study of history was persistently pursued.

Cin Condition:

1. Now maybe they'd realize that life can be tough.
2. His work began just six days after the flood.
3. No one will even suspect that it is your work.
4. Yet he remains the fiercest of competitors.
5. You're so preoccupied that you've let your faith grow dim.
6. He swung up over the wheel.
7. The buddies invariably took advantage of him.
8. But it was a hopeful sign, he told himself.
9. Ralph prepared red snapper with fresh lemon sauce for dinner.
10. The blow encountered silky hair and hard bone.
11. The fear of punishment just didn't bother him.
12. Instead of that he was engulfed by bedlam.
13. Do you always navigate like this?
14. But that explanation is only partly true.

REFERENCES

American National Standards Institute. (1969) *American National Standards for Calculation or the Articulation Index* (ANSI S3.5-1969). New York: Author.

American National Standards Institute. (1996) Specification *for Audiometers* (ANSI S3.6-1996). New York: Author.

Baer, T., and Moore, B. C. J. (1993). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. 94, 1229-1241.

Balakrishnan, U., Freyman, R. L., Chiang, Y., Nerbonne, P. G., and Shea, K. (1996). "Consonant recognition for spectrally degraded speech as a function of consonant-vowel intensity ratio," J. Acoust. Soc. Am. 99 (6), 3758-3768.

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., & Rzeczkowski, C. (1984) Standardization of a test of speech perception in noise. Journal of Speech and Hearing Research, 27, 32-48.

Caramazza, A., Chialant, D., Capasso, R., and Miceli G. (2000). Separable processing of consonants and vowels. Nature 403, 159-160.

Cole, R., Yan, Y., Mak, B., Fanty, M., and Bailey, T. (1996). Journal Proceedings of the 1996 International Conference on Acoustics, Speech, and Signal Processing. Atlanta, Ga., May 1996.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., and Gerstman, L. J. (1952). "Some experiments on the perception of synthetic speech sounds," J. Acoust. Soc. Am. 24, 597-606.

Dunn, H. K. and White, S. D. (1940). "Statistical measurements on conversational speech," J. Acoust. Soc. Am. 11, 278-288.

Egan, J. (1948). Articulation testing methods. *The Laryngoscope*, 58, 955-991.

Era, P., Jokela, J., Qvarnberg, Y., and Heikkinen, E. (1986). “Pure-tone thresholds, speech understanding, and their correlates in samples of men of different ages,” Audiology 25, 338-352.

Faulkner, A., and Rosen, S. (1999). “Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech”, J. Acoust. Soc. Am. 106 (4), Pt. 1, 2063-2073.

Fletcher, H., *Speech and Hearing*. New York: Van Nostrand (1929).

Folstein MF, Folstein SE, McHugh PR. (1975) “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 12 (1): 59-68.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (1993). DARPA TIMIT: Acoustic-Phonetic Continuous Speech Corpus.

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

Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, G., Eldert, E., and Benson, R. W. (1952). Development of materials for speech audiometry. Journal of Speech & Hearing Disorders, 17, 321-337.

Hultzen, Lee S., Allen, Joseph H.D., Jr., and Miron, Murray S. (1964). “Tables of Transitional Frequencies of English Phonemes”, University of Illinois Press, Urbana.

Kalikow, D.N., Stevens, K.N., & Elliot, L.L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J. Acoust. Soc. Am., 61, 1337-1351.

Kuchera, Henry and Francis, W. Nelson (1967). "Computational Analysis of Present-Day American English," Brown University Press, Providence, Rhode Island.

Liberman, A.M., Cooper, F.S., Shankweiler, D.B., and Studdert-Kennedy, M. (1967). Perception of the speech code, Psych. Rev. 74, 431-461.

Miller, George A., (1951) Language and Communication. New York, NY: McGraw-Hill Book Company, Inc.

Musiek, F., Rintelmann, W. (1999) Contemporary Perspectives in Hearing Assessment. Needham Heights, MA: Allyn & Bacon, Publishers.

Owens, E., Talbott, C., Schubert, E. (1968) Vowel Discrimination of Hearing-Impaired Listeners. Journal of Speech and Hearing Research, 11 (3), 648-655.

Oxenham, Andrew J., Bacon, Sid P. (2003). "Cochlear Compression: Perceptual Measures and Implications for Normal and Impaired Hearing," Ear & Hearing, 24, 352-366.

Plomp, R. (2002). The Intelligent Ear: On the Nature of Sound Perception. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Plomp, R. (1983). The role of modulation in hearing. In R. Klinke & physics (pp. 270-276). Berlin: Springer-Verlag.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," Science 270, 303-304.

Studebaker, G.A. (1985). A “rationalized” arcsine transform. *J Speech Hear Res, 28,* 455-462.

Van Tasell, D.J. (1993). Hearing loss speech and hearing aids. J. Speech Hear. Res., 36, 228-244.

Wingfield, A., and Tun, P.A. (2001). “Spoken language comprehension in older adults: Interactions between sensory and cognitive change in normal aging. Seminars in Hearing, 22, 287-301.

CURRICULUM VITAE

THOMAS Z. BURKLE

IDENTIFICATION
Birth: May 14, 1975
Location: Indianapolis, IN

ADDRESS
2214 S Oakdale West Dr
Bloomington, IN 47403

EDUCATION

2002-Present	Indiana University Bloomington, Indiana Major: Audiology	Au.D. candidate
2002-2003	Indiana University Bloomington, Indiana Major: Audiology Degree Awarded: December 2003	M.A.
1997-2001	Ball State University Muncie, Indiana Major: Music Engineering Technology Degree Awarded: June 2001	B.M.

HONORS AND AWARDS

Outstanding Graduate Clinician Awarded by ISHA, 2004
Northside Exchange Club Scholarship, 2003-2004

PROFESSIONAL EXPERIENCE

Clinical Audiologist Trainee, Richard L. Roudebush Veterans Administration Medical Center, Indianapolis, IN
February 2004-Present

Research Assistant, Indiana University Speech Psychoacoustics Laboratory, Bloomington, IN
January 2003-Present

PUBLICATIONS

Kewley-Port, D., Liu, C. and Burkle, T. Zachary. (2003) Vowel formant discrimination in high-fidelity speech by hearing-impaired listeners. Presented at the 146th meeting of the Acoustical Society of America, Austin, TX, Nov., 2003.