J Am Acad Audiol 5: 255–258 (1994)

Word Recognition with Segmented-Alternated CVC Words: Compact Disc Trials

Richard H. Wilson*

Abstract

The *Tonal and Speech Materials for Auditory Perceptual Assessment, Disc 1.0* audio compact disc contains a set of consonant, vowel-nucleus, consonant speech materials that have been segmented at the approximate phoneme boundaries so that the carrier phrase and consonant segments of a word can be presented to one ear and the carrier phrase and vowel segment of the word presented to the other ear. In this manner, consonant and vowel information are presented to different ears using a binaural fusion paradigm in which the temporal sequence of the word is maintained. The purpose of the compact disc trials was to document the recognition performance of 120 young adults with normal hearing under the following three conditions: (1) the consonant segments presented monaurally at 40, 50, and 60 dB HL; (2) the vowel segment of the words presented monaurally at 20, 30, and 40 dB HL; and (3) the binaural condition at 0 to 30 dB HL in which the consonant segments were presented to one ear and the vowel segments were presented to the other ear. Recognition performance under the binaural fusion paradigm was identical to recognition performance on the same materials unaltered.

Key Words: Binaural temporal fusion, compact disc, CVC words, segmented words

This paper describes a binaural temporal fusion paradigm for word recognition that is included on the *Tonal and Speech Materials for Auditory Perceptual Assessment, Disc 1.0* compact disc (Wilson et al, 1984). This paradigm, which uses the consonant-vowel nucleus-consonant (CVC) words recorded with the Picture Identification Task materials (Wilson and Antablin, 1980, 1982), involves presentation of the consonant segments of the word to one ear and presentation of the vowel segment of the word to the other ear. In this segmented-alternated binaural paradigm, the temporal sequence of the word is maintained.

Two binaural fusion paradigms have been used to assess the integrity of the central auditory system. *Binaural temporal fusion* paradigms typically involve a speech signal that is alternated with a 50 percent duty cycle up to 100 times/sec between ears. In addition to the degradation of the speech signal caused by the alterations, the distortion caused by the rapid switching introduces further degradation of the signal. Some patients with neurologic lesions demonstrate poor performance on this type of binaural temporal fusion task (Bocca and Calearo, 1963; Lynn and Gilroy, 1977). *Binaural spectral fusion* paradigms typically involve filtering a speech signal into a low-pass band and a high-pass band (Matzker, 1959). The goal with these schemes was to present minimal information to each ear that when combined (fused) produced recognition performance that was substantially more than the sum of the recognition performances on the two parts. Smith and Resnick (1972) used a 360–890 Hz pass band and a 1750–2220 Hz pass band. By design, the binaural spectral fusion paradigm simultaneously placed predominantly vowel information in one ear (low-pass band) and predominantly consonant information in the other ear (high-pass band). When either pass band was presented alone, word-recognition performance for the listener with normal hearing was about 20 percent correct. When presented in the binaural paradigm with the low-pass signal to one ear and the high-pass signal to the other ear, recognition performance improved to about 60 percent correct. Because substantial portions

*VA Medical Center, Mountain Home, Tennessee, East Tennessee State University, Johnson City, Tennessee

Reprint requests: Richard H. Wilson, Audiology — 126, VA Medical Center, Mountain Home, TN 37684

of the speech spectrum had to be eliminated in these binaural spectral fusion tasks, the materials were difficult to use with subjects with hearing loss.

Ten years ago, our laboratory embarked on a project, the goal of which was "... to develop a set of speech stimuli that are sensitive to central auditory disorders but are relatively immune to the effects of peripheral hearing loss" (Wilson et al, 1984, pp. 378–379). We reasoned that, "if words are segmented to present essentially vowel information to one ear and consonant information to the other ear without reducing the spectral or temporal information, then peripheral hearing loss should not substantially reduce performance on segmented-alternated words with reference to word-recognition performance on traditional monosyllabic word lists" (p. 379). To achieve this goal, consonant-vowel-consonant monosyllabic words were segmented into approximate consonant phoneme and approximate vowel phoneme units "in which the majority of acoustic and linguistic cues are restricted to the appropriate consonant segment or vowel segment" (p. 379).

An example digital waveform of the speech signals that were developed is given in Figure 1. The carrier phrase is "Show me," and the target word is "yam." In the example, the carrier phrase and the vowel segment of the word are on channel 1, and the carrier phrase and the consonant segments of the word are on channel 2. A waveform editor was used to segment the words, with listener judgments providing the final decision. Clicks at the edited segment boundaries were avoided by shaping the transitions, "with a ¼-cycle cosine ramp during the first and last 10 msec of each speech sound segment. For onset shaping, the function argument was incremented from $\pi/2$ to π radians, whereas for offset shaping the function argument was incremented from 0 to $\pi/2\pi$ radians" (p. 380). The 200 edited words were output (Nicolet, Model 812) by two 12-bit digital-to-analog converters and recorded on ½-track analog tape (Revox, B77). To avoid a midline image, the carrier phrases on the two channels were recorded 180 degrees out of phase. After a series of four experiments, the segmented-alternated words were reduced to 50 select words that produced about 10 percent correct word recognition in response to monaural presentation of the vowel segments (at 50 dB SPL) or the consonant segments (at 70 dB SPL). Two randomizations of these 50 words were included on the *Tonal and Speech Materials for Auditory Perceptual Assessment* audio compact disc. During the compact disc trials, the word-recognition performance for the monaurally presented consonant segments and the monaurally presented vowel segment of the select 50 words was evaluated over a more extensive range of sound pressure levels than were included in the initial investigation.

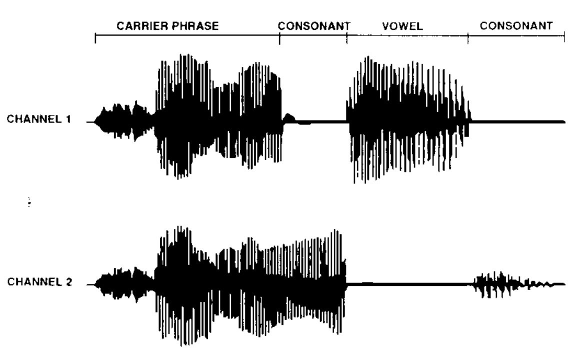

Figure 1 Amplitude (Y) by time (X) display of the digitized waveform "Show me yam" that illustrates the segmented-alternated word paradigm in which the carrier phrase and vowel segment of the CVC word are presented on channel 1 and the carrier phrase and the consonant segments of the CVC word are presented on channel 2.

METHOD

The 50 segmented-alternated CVC words were digitized from the master analog recording (Revox, Model B77) using a 16-bit analog-to-digital converter (Antex, Model SX-10) at 20,000 samples/sec and an antialaising filter set to 8800 Hz (96 dB/octave rejection). The files were edited to minimize the length. (These are the digital masters from which the Picture Identification Task materials included on the *Speech Recognition and Identification Materials, Disc 1.1* compact disc were made.)

Three conditions (tracks 9 and 10) were studied in the compact disc trials with 120 subjects with normal hearing (Noffsinger et al, 1994). First, the carrier phrase and the consonant segments of the words were presented monaurally at 40, 50, and 60 dB HL (60 to 80 dB SPL). Second, the carrier phrase and the vowel segment of the words were presented monaurally at 20, 30, and 40 dB HL (40 to 60 dB SPL). Third, a psychometric function from 0 to 30 dB HL (20 to 50 dB SPL) in 6-dB steps was generated for the binaural condition in which the vowel segments were presented to one ear and the consonant segments were presented to the other ear. The levels of the two components in the binaural condition were unaltered with respect to the VU calibration of the nonsegmented materials. The two monaural conditions were presented before the binaural condition. To minimize learning effects, each subject only listened to one level of each of the three conditions. Thus, each mean datum point represents data from 20 different subjects for the binaural conditions and data from 40 different subjects for the monaural conditions.

Following completion of the clinical trials with 120 subjects, it was apparent that recognition performance at the highest presentation levels was not asymptotic for the consonant segments of the words and for the vowel segments of the words presented monaurally. Supplemental data for the monaural consonant and vowel segments (tracks 9 and 10), therefore, were collected at 55 dB HL and 70 dB HL from 40 subjects (mean = 27.6 years) with normal hearing who were new to the study.

Table 1 Mean and Standard Deviation Percent Correct Recognition at Respective Presentation Levels for Consonant Segment Materials, Vowel Segment Materials, and Materials Presented Binaurally

	Presentation Level (dB HL)				
Monaural Presentation	*20*	*30*	*40*	*50*	*60*
Consonant Segments					
Mean			17.0	34.6	66.8
SD			10.1	19.8	25.6
Vowel Segments					
Mean	12.8	12.2	20.4		
SD	9.9	7.7	19.9		

	Presentation Level (dB HL)					
*Binaural Presentation**	*0*	*6*	*12*	*18*	*24*	*30*
Mean	11.1	41.0	70.0	78.6	95.8	95.2
SD	11.8	17.3	12.3	13.2	3.5	3.9

Each datum point for the monaural presentations (top) was obtained from 40 listeners with normal hearing, whereas each datum point for the binaural presentation (bottom) was obtained from 20 listeners with normal hearing.

*Vowels in one ear, consonants in the other.

RESULTS AND DISCUSSION

The data from the compact disc trials are depicted in Figure 2 and listed in Table 1. In the figure, the percent correct recognition obtained from the compact disc trials is displayed for the binaural condition in which the vowel segments were presented to one ear and the consonant segments were presented to the other ear (open circles) for the vowel segment condition (squares) and for the consonant segment condition (triangles). The means for the supplemental monaural data for the consonant segments (filled triangles) and for the vowel segments (filled squares) also are depicted in Figure 2. The normal, monaural psychometric function for the Picture Identification Task materials is shown for reference (filled circles) (from Wilson and Antablin, 1980). Several features are noteworthy from the data in the figure. First, the psychometric function for the compact disc trials is essentially identical with the psychometric function for the normative, monaural data for the materials. The slope of the binaural function[1] is 4.0 percent/dB between the 20 and 80 percent correct points and is not different from the 4.1 percent/dB reported with the original monaural function.

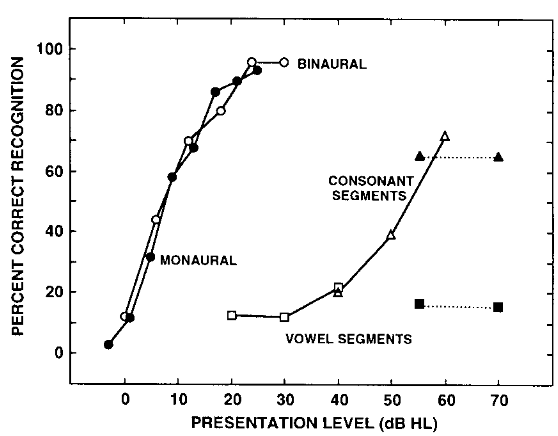

Figure 2 Percent correct recognition obtained from the compact disc trials for the binaural condition (20 subjects/datum point) in which the vowel segments were presented to one ear and the consonant segments were presented to the other ear (open circles) for two monaural conditions (40 subjects/datum point): vowel segments (squares) and consonant segments (triangles). Filled symbols represent data from the 40 subjects in the supplemental study. Normal, monaural psychometric function for the Picture Identification Task materials is shown for reference (filled circles) (from Wilson and Antablin, 1980).

[1]Mean data for the three experimental conditions were fit with second- or third-degree polynomials from which the displacements of the functions and the slopes of the functions were calculated. The following equations were generated to represent the functions (R^2 = 0.99 or 1.00):

consonant segments: $y = 92.600 - 4.8100x + 0.07300x^2$;
vowel segments: $y = 40.400 - 2.2600x + 0.04400x^2$;
binaural: $y = 10.934 + 5.8508x - 0.10388x^2 + 0.0001072x^3$;
and
monaural: $y = 11.121 + 3.8575x + 0.14081x^2 - 0.0066525x^3$.

With the monaural vowel segments, there was only minor improvement (8%) as the presentation level was increased from 20 to 40 dB HL. Somewhat unexpectedly, the recognition performance for the monaural consonant segments condition improved substantially as the presentation level increased from 40 to 60 dB HL. Over this 20-dB range, the mean recognition performance improved from 17 percent correct to 67 percent, which is the increase in performance that prompted the supplemental study. The clinical trials data and the supplemental data (filled symbols) suggest: (1) that the function for the monaural consonant segments becomes asymptotic around 65–70 percent correct over the 55 to 70 dB HL range; and (2) that the function for the monaural vowel segments is constant at 15–20 percent correct over the 20 to 70 dB HL range.

In all probability, the improvement in recognition performance on the monaural consonant condition with increasing presentation level is owing to the relation between the presentation level and the level of energy in the consonant segments of the words. As illustrated by Bornstein et al (1994, Fig. 1), the low-frequency energy in the speech spectrum (vowels) peaks at 0 dB, whereas the high-frequency energy in the speech spectrum (consonants) peaks at –30 dB. As the consonant segments are presented at higher levels, the segments become more audible, and the listener is able to obtain more information from the signal. This improvement in recognition performance with the consonant segments was not observed in the earlier study (Wilson et al, 1984) because only one level was investigated, 50 dB HL. Further, when the materials were originally devised, the final judge of the success of segmentation was based on perceptual evaluations made at 70 dB SPL, which was thought to be at a level high enough to preclude improved performance. In retrospect, this was an error in judgment. It remains impressive, however, that at a level of 40 dB HL, at which recognition performance on either of the monaural segments was about 20 percent correct, binaural performance was substantially higher, more than simply an additive effect. In this example, at 40 dB HL, binaural performance was about 100 percent correct or 80 percent better than performance under either monaural condition. In all probability, at presentation levels below 40 dB HL, the difference between performances under the monaural and binaural conditions would be extended. The effects of this binaural temporal fusion paradigm need to be defined on various patient populations, especially patients who have a variety of neurologic impairments that range from the subtle involvements encountered in the normal aging process to the involvements associated with the various space-occupying lesions.

Acknowledgment. This work was supported by Medical Research Service and Rehabilitation Research and Development Service, Department of Veterans Affairs.

REFERENCES

Bocca E, Calearo C. (1963). Central hearing processes. In: Jerger J, ed. *Modern Developments in Audiology.* New York: Academic Press, 337–370.

Bornstein SP, Wilson RH, Cambron NK. (1994). Low- and high-pass filtered Northwestern University Auditory Test No. 6 for monaural and binaural evaluation. *J Am Acad Audiol* 5:259–264.

Lynn GE, Gilroy J. (1977). Evaluation of central auditory dysfunction in patients with neurological disorders. In: Keith RW, ed. *Central Auditory Dysfunction.* New York: Grune and Stratton, 177–221.

Matzker J. (1959). Two new methods for the assessment of central auditory functions in cases of brain disease. *Ann Otol Rhinol Laryngol* 68:1185–1197.

Noffsinger D, Musiek FE, Wilson RH. (1994). Department of Veterans Affairs compact disc recording for auditory perceptual assessment: background and introduction. *J Am Acad Audiol* 5:231–235.

Smith BB, Resnick DM. (1972). An auditory test for assessing brain stem integrity: preliminary report. *Laryngoscope* 82:414–424.

Wilson RH, Antablin JK. (1980). A Picture Identification Task as an estimate of the word-recognition performance of nonverbal adults. *J Speech Hear Disord* 45:223–238.

Wilson RH, Antablin JK. (1982). The Picture Identification Task — a reply to Dillon. *J Speech Hear Disord* 47:111–112.

Wilson RH, Arcos JT, Jones HC. (1984). Word recognition with segmented-alternated CVC words: a preliminary report on listeners with normal hearing. *J Speech Hear Res* 27:378–386.