Interpreting word-recognition data using the NAM and phonemic features

RACHEL MCARDLE[1] AND RICHARD H. WILSON[2]

[1] *Auditory Research Laboratory (126), Bay Pines VA Healthcare System, Bay Pines, Florida, 33744, USA and Department of Communicative Disorders and Sciences, University of South Florida, Tampa, Florida, 33620, USA*

[2] *Auditory Research Laboratory (126), James H. Quillen VA Medical Center, Mountain Home, Tennessee, 37684, USA and Departments of Surgery and Communicative Disorders, East Tennessee State University, Johnson City, Tennessee, 37614, USA*

The goal of this project was to examine acoustic and non-acoustic variables that may predict the relative ease or difficulty with which monosyllabic words presented in speech-spectrum noise are recognized. For the analysis, the 50% correct recognition data from the 24 listeners with normal hearing who participated in the Wilson and McArdle (2007) was used. The following acoustic, phonetic/phonological, and lexical variables were included in the evaluation: (1) rms; (2) duration; (3) consonant features (manner, place, and voicing for initial and final phoneme); (4) vowel phoneme; (5) word frequency; (6) word familiarity; (7) neighborhood density; and (8) neighborhood frequency. The results showed significant correlations between the acoustic variables (i.e., rms, duration) and the 50% point. The results of the regression analysis found that 45% of the variance associated with the 50% point was accounted for by the acoustic and phonetic/phonological variables (i.e., consonant features, vowel phoneme) whereas only 3% of the variance was accounted for by a single lexical variable (i.e., word familiarity). Word frequency, neighborhood density, and neighborhood frequency were not found to be significant variables in the regression model. These findings suggest that monosyllabic word-recognition-in-noise is more dependent on bottom-up processing than top-down processing. Thus, monosyllabic words may be more sensitive to changes in audibility when using speech-in-noise testing for rehabilitative outcomes such as in a pre/post-hearing aid fitting format.

INTRODUCTION

Several speech-in-noise tests have been developed for research and clinical use, the majority of which use sentence materials, however, the performance on sentence measures can be influenced by syntactic and semantic context. The influence of context on performance may be representative of how an individual communicates in everyday life but the basic auditory function on an individual with speech signals at the sensory level may be masked by the involvement of higher-level information.

A major role of clinical speech-in-noise testing is to use the individual recognition performance results to optimize the fitting of sensory aids. With that purpose then,

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.).
ISBN: 87-990013-1-4. Print: Centertryk A/S.

one would want to minimize top-down influences such as semantic and syntactic context on performance. Both word and sentence materials activate both bottom-up (i.e., incoming auditory information) and top-down processing during a recognition task, however, the amount of bottom-up and top-down processing is inversely related for word and sentence materials. Although the use of monosyllable words is thought to minimize the influence of higher level processing such as semantic and syntactic contextual cues, past studies, including those examining the theory of the Neighborhood Activation Model (NAM), suggest that recognition performance for isolated words also is influenced by top-down lexical processing (Luce and Pisoni, 1998).

The purpose of this study was to analyze the 50% correct recognition data from Wilson and McArdle (2007) to examine acoustic, phonetic/phonological, and lexical variables that may predict the relative ease or difficulty with which monosyllables presented in noise are recognized. The following variables were included: (1) rms; (2) duration; (3) phonetic content (initial and final consonant manner, place, and voicing); (4) vowel phoneme; (5) word frequency; (6) word familiarity; (7) neighborhood density; and (8) neighborhood frequency.

METHODS

Psychometric data

The psychometric data used in the current study was obtained from Wilson and McArdle (2007). Specifically, the 50% points were used from the 490 monosyllablic words that were presented to 24 listeners with normal hearing at 4 signal-to-noise ratios. The monosyllabic words were generated from four lists of each of the following materials: PB-50, CID W-22, and NU No. 6. The monosyllabic words were interspersed and recorded by a single female speaker. Percent correct values were obtained for each word at each of four signal-to-noise ratios (SNR, S/N) and the 50% point was established by using the Spearman-Kärber equation (Finney, 1952; Wilson *et al*, 1973).

Procedures

To obtain measurements for the acoustic variable, sound editing software (Adobe Audition 2.0) was used to measure the rms and duration of each of the monosyllabic words. For the phonetic/phonological variables, including manner, place, and voicing of the initial and final phonemes and the vowels, sound files from the recorded materials were transcribed by a phonologist. For the lexical variables, such as word frequency, familiarity, neighborhood density, and neighborhood frequency the individual values for each word were obtained from the 20,000-word Hoosier Mental Lexicon (Nusbaum *et al*, 1984). Initially 547 monosyllabic words were reported by Wilson and McArdle (2007), however, 57 monosyllabic words were not listed in the Hoosier Mental Lexicon, therefore reducing the corpus of words used in the current analysis to 490.

In addition to the absolute values obtained from the Hoosier Mental Lexicon for the lexical variables, median splits were performed for word frequency, neighborhood density, and neighborhood frequency in order to categorize the words as high (H) or low (L) with respect to the median. This procedure has been used in previous studies (e.g., Sommers, 1996; Dirks *et al*, 2001).

RESULTS

Acoustic variables

To examine the relationship between rms and the 50% point, a Pearson Product-Moment correlation was used. The result was statistically significant (r = 0.16, p<.01) but the size of the correlation suggests little if any relationship between rms and recognition performance. Similarly, a Pearson Product-Moment correlation was used to examine the relationship between duration and the 50% point. The result showed a statistically significant correlation (*r* = -0.18, *p*<.01) but again, the size of the correlation suggests little if any relationship between duration and recognition performance. The slight relationship shows that shorter words required a more positive SNR than the longer words at the 50% point. As with the rms data, the large data set size probably made the Pearson Product-Moment correlation overly sensitive. The slight relationship found between rms and duration was somewhat expected given that only monosyllabic words were recorded at similar levels in conjunction with a carrier phrase to reduce inter-stimulus variability for the initial experiment (Wilson *et al*, 2007). The range in terms of duration for the monosyllabic words, however, was wide in that the longest word (750 ms) was almost three times the length of the shortest word (240 ms). The same was true for rms with the weakest word (-25.8 dB, re: maximum digitization range) 9 dB below the strongest word (-16.8 dB).

Phonetic/Phonological variables

Consonant Features. Table 1 lists the mean 50% correct recognition points (and standard deviations) for the individual words as a function of consonant features such as manner, place, and voicing of the initial and final phoneme speech sounds. The following observations can be seen in the table for each consonant feature:

(1) *Manner* – The 50% points ranged from the most difficult to understand in noise, liquids (/l,r/), with a mean 50% point of 2.8-dB S/N and 2.5-dB S/N for the initial and final positions, respectively, to the easiest to understand in noise, affricates (i.e., /dʒ, tʃ/), with a mean 50% point of -1.9-dB S/N and -2.9-dB S/N for the initial and final positions. The dB range for 50% points from the easiest to the most difficult manner was 4.7 dB and 5.4 dB for initial and final position, respectively.

(2) *Place* - The dB range of 50% points from the easiest to the most difficult place was 4.1 dB and 4.8 dB for initial and final position, respectively. Bilabials (/b,p,m/) were the most difficult to understand in noise, whereas the alveopalatals (i.e., /dʒ, tʃ, ʃ, ʒ/) were the easiest to understand. The phonemes (i.e., /dʒ, tʃ/), which belong to the affricate manner and the alveopalatal place categories, were the easiest to recognize in noise for both initial and final position, making it unclear as to whether it is the consonant feature of manner or place that positively affected recognition performance.

(3) *Voicing* - The voiceless phonemes in the initial (-0.2-dB S/N) and final (-0.7-dB S/N) positions were recognized at lower SNRs than were the voiced phonemes in the initial (1.6 dB S/N) and final (1.9-dB S/N) positions. This finding is consistent with Miller and Nicely (1955) who reported that voiceless phonemes in English are more intense given that voiceless phonemes are aperiodic and noisier than voiced phonemes.

	Phoneme Position							
	Initial	Final	Initial	Final	Initial	Final	Initial	Final
MANNER								
	Stops		Fricatives		Affricates		Nasals	
Mean	1.0	0.8	-0.1	-0.6	-1.9	-2.9	0.9	1.8
SD	3.3	3.5	3.3	3.0	4.5	2.9	3.2	2.8
	Liquids		Glides					
Mean	2.8	2.5	0.6	--				
SD	2.6	3.0	2.3	--				
PLACE								
	Bilabials		Labiodental		Dental		Alveolar	
Mean	2.4	2.2	0.3	0.8	1.6	-0.5	0.7	0.6
SD	3.2	3.2	2.6	2.9	2.3	2.6	3.5	3.4
	Alveopalatal		Velar		Labio-velar		Glottal	
Mean	-1.7	-2.6	0.7	1.1	0.8	--	0.9	--
SD	3.8	2.6	3.0	3.4	2.4	--	3.3	--
VOICING								
	Voiceless		Voiced					
Mean	-0.2	-0.7	1.6	1.9				
SD	3.3	3.1	3.2	3.2				

Table 1: Mean 50% points (dB S/N) and standard deviations (dB) are listed for the consonant features of manner, place, and voicing of the initial and final phonemes.

Vowel phonemes. Table 2 lists the mean 50% points (dB S/N), and the standard deviations (dB) for the 18 vowel phonemes. The vowel phonemes are listed from easiest to understand in noise (i.e., /ju/, -2.2-dB S/N) to most difficult to understand in noise (/ɑ/, 1.9-dB S/N), a 4.1 dB range. A one-way analysis of variance found no significant difference [F(17,489) = 1.45, p > .05] in mean 50% points across the 18 vowel phoneme categories. Although the three vowel phonemes associated with the easiest word recognition in noise as measured by the mean 50% point for the whole words were diphthongs (i.e., /ju, ɛr, aʊ/) and the three vowel phonemes associated with the most difficult recognition in noise as measured by the mean 50% point for the whole words were monophthongs (i.e., /ɔ, ɪ, ɑ/), there was no significant difference [F(1,489) = 0.11, p

> .05] in mean 50% points for the monosyllabic words with a monophthong (n = 339) or a diphthong (n = 151) as the vowel phoneme.

Vowel Symbol	Examples	N	50% pt	SD
/ju/	cued, you, juice	9	-2.2	3.2
/ɛr/	hair, there, tare	4	-1.3	2.8
/aʊ/	how, shall, doubt	13	-0.8	4.4
/ɝ/	hurt, search, third	18	-0.2	2.9
/ɑr/	hard, carve, bar	16	-0.1	4.0
/oʊ/	code, foe, though	34	0.0	4.6
/u/	hoot, goose, chew	22	0.2	3.1
/ae /	had, bath, and	60	0.3	3.6
/ɛ/	head, chess, said	35	0.3	3.2
/eɪ/	hate, chain, drake	39	0.4	3.4
/ɔɪ/	boyd, void, toy	6	0.6	3.0
/ʊ/	hood, cook, should	11	1.1	2.1
/ʌ/	bud, tub, scrub	47	1.1	3.8
/i/	heat, keen, seize	34	1.2	3.4
/aɪ/	high, rhyme, wife	35	1.2	3.2
/ɔ/, /ɔr/	jaw, flaunt, hog	40	1.4	3.4
/ɪ/, /ɪr/	hit, bliss, gin	61	1.4	3.0
/ɑ/	hot, cod, wash	6	1.9	4.1

Table 2: The mean 50% correct recognition points (dB S/N) and standard deviations (dB) for the words that contained the respective vowel phonemes.

Lexical variables

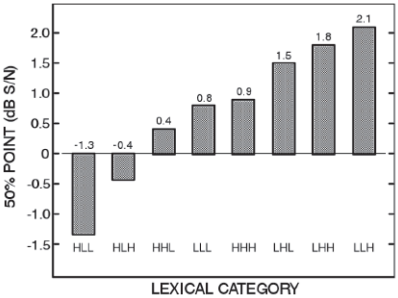

Fig. 1: Means 50% points (ordinate) for each exical category (abscissa).All lexical categories are nemed uæsing the same convention such that target word frequency is indicated as low (L) or high (H) in the first position, neighborhood density is indicated as low (L) or high (H) in the second position, and neighborhood frequency is indicated as low (L) or high (H) in the third position.

The words in the current study were categorized by word frequency followed by neighborhood density and then neighborhood frequency such that words in the HLL group had High word frequency, Low-neighborhood density, and Low-neighborhood frequency. Figure 1 provides the mean 50% correct recognition points (dB S/N) for the words in each of the 8 NAM categories. As shown in Figure 1, the HLL words required the lowest SNR (-1.3 dB) to obtain 50% correct recognition, whereas the LLH words required the highest SNR (2.1 dB) to obtain 50% correct. A one-way analysis of variance revealed a significant effect of NAM category [$F(7,489) = 6.9, p<.001$]. Post hoc comparisons using a Bonferroni correction for multiple t-tests showed that the easiest words, the HLL words, were not significantly different than the HLH or the HHL words, which were the next two easiest categories in terms of recognition performance, but were significantly different than all five other categories. The two most difficult categories in terms of recognition performance, the LLH and the LHH words, were only significantly different from the two easiest categories, the HLL and the HLH group.

Regression Analysis

The data for all 490 monosyllable words were analyzed using multiple linear regressions to identify predictor variables that best influenced the 50% point. The predictor variables included in the regression analysis were: rms; duration; initial phoneme manner, place, and voicing; final phoneme manner, place, and voicing; vowel phoneme; word frequency; word familiarity; neighborhood density; and neighborhood frequency. All categorical variables were dummy-coded1 for use in the regression analysis. The decibel SNR for the mean 50% point for the 24 listeners was used as the criterion variable. A significant model emerged [$F(28,461) = 17.41, p < .001$, adjusted $R^2 = 0.48$] (see Table 3). Three predictor variables (vowel phoneme, neighborhood density, and neighborhood frequency) did not add significantly to the regression model and are not listed.

Predictor Variable	R^2	Adjusted R^2	R^2 Change
rms	0.04	0.04	0.04
Duration	0.06	0.05	0.02
Initial Phoneme Manner	0.16	0.15	0.10
Initial Phoneme Place	0.22	0.20	0.06
Initial Phoneme Voicing	0.23	0.20	0.01
Final Phoneme Manner	0.37	0.34	0.14
Final Phoneme Place	0.41	0.38	0.04
Final Phoneme Voicing	0.48	0.45	0.07
Familiarity	0.51	0.48	0.03

Table 3: Significant predictor variables from the linear regression are listed in column 1. The R^2 values, and the adjusted R^2 values in the 2nd and 3rd columns, respectively show the amount of variance accounted for by the addition of each variable such that the total variance accounted for by the model is shown on the last row of the table. The R^2 change values in the fourth column show the variance accounted for by each individual variable.

For each predictor variable the R^2 value, adjusted R^2 value, and the change in R^2 are reported in Table 3. Interestingly, the majority of the variance was accounted for by the consonant features of the initial and final phoneme for which the sum of the R^2 change values was 42%. Familiarity was the only lexical variable to account for a significant amount of variance in relation to the 50% point and resulted in a change in R2 of 3%.

CONCLUSIONS

The following conclusions can be drawn from the data in the present study when young listeners with normal hearing are presented materials in speech-spectrum noise that are spoken on one occasion by a single speaker:

(1) Acoustic and phonetic/phonological variables associated with bottom-up processing (i.e., rms, duration, articulatory characteristics of the consonants in initial and final position) can predict almost half of the variance associated with the recognition performance at the 50% point.

(2) With the exception of word familiarity, other lexical variables associated with top-down processing (i.e., word frequency, neighborhood density, neighborhood frequency) were not found to be significant predictors of word-recognition performance at the 50% point.

(3) If using speech-in-noise testing in a pre- and post-hearing aid fitting format, then the use of monosyllabic words may be more sensitive to changes in audibility resulting from the hearing aids than would be contextual sentence materials.

[1] Categorical variables that have more than two levels are dummy-coded with ones and zeros to identify category membership.

ACKNOWLEDGMENT

The Rehabilitation Research and Development Service, Department of Veterans Affairs supported this work through a Research Career Development award to the first author and a Senior Research Career Scientist award and a Merit Review for the Auditory and Vestibular Dysfunction Research Enhancement Award Program (REAP) to the second author. A more complete version of this report will appear in a forthcoming issue of the Journal of the American Academy of Audiology.

REFERENCES

Dirks, D. D., Takayanagi, S., Moshfegh, P., Noffsinger, P. D., and Fausti, S.A. (**2001**). “Examination of the neighborhood activation theory in normal and hearing-impaired listeners,” Ear Hear. **2,** 1-13.

Finney, D. J. (**1952**). “Statistical method in biological assay,” London: C. Griffen.

Luce, P. A. and Pisoni, D. B. (**1998**). “Recognizing spoken words: The neighborhood activation model,” Ear Hear. **19**, 1-36.

Miller, G. A., and Nicely, P. E. (**1955**). “An analysis of perceptual confusions among some English consonants,” J. Acoust. Soc. Am. **27**, 338-352.

Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (**1984**). “Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words,” Research on Speech Perception Progress Report No. 10. Bloomington, IN: Speech Research laboratory, Psychology Department, Indiana University.

Sommers, M. (**1996**). The structural organization of the mental lexicon and its contribution to age-related declines in spoken word recognition. Psychology and Aging, **11**:333-341.

Wilson, R.H., Morgan, D.E., and Dirks, D.D. (**1973**). A proposed SRT procedure and its statistical precedent. J Speech Hear Disord, **38**:184-191.

Wilson, R. H., and McArdle, R. A. "Recognition Performance on Single-speaker Recordings of W-22, NU6, & PB-50 by Listeners with Normal Hearing." International Symposium on Auditory and Audiological Research, Helsingør, Denmark (**2007**).