List Equivalency of the Northwestern University Auditory Test No. 6 in Quiet and in Continuous Broad Band Noise

Equivalence des listes de l'épreuve auditive no. 6 de l'Université Northwestern administrée dans le silence et dans un bruit à large bande continu

Andrew Stuart, MSc
Department of Psychology
Dalhousie University

Walter B. Green, PhD
School of Human Communication Disorders
Dalhousie University

Dennis P. Phillips, PhD
Department of Psychology
Dalhousie University

Robert Stenstrom, MA
Dept. of Epidemiology and Biostatics
McGill University

Key words: word recognition, speech in noise, list equivalency, Northwestern University Test No. 6

Abstract

Interlist equivalency of lists one to four of the *Northwestern University Auditory Test No. 6*, in quiet and in continuous broad band noise, was investigated with 48 normal hearing young adults. All lists were administered at 50 dB sensation level in quiet and in noise at signal-to-noise ratios (S/Ns) of 10, 5, 0, -5, -10, -15, and -20 dB. Results indicated no statistical differences in word recognition performance among lists when administrated in quiet (p = .542). However, a significant list effect was found when administered in noise (p = .0082). In general, such differences are taken to support the caution that interpretation of performance differences of subjects on speech tasks should be tempered when presented in a background of noise. In practice, however, the size of the differences was so small (e.g., less than 1% of data variance was accounted for, and mean list differences never exceeded 7.9% or were less than four words) that the effect was seen as behaviorally/clinically irrelevant.

Abrégé

En vue de déterminer l'équivalence entre les listes un à quatre de l'épreuve auditive no 6 de l'Université Northwestern, on a fait subir l'épreuve, dans le silence et dans un bruit à large bande continu, à 18 jeunes adultes entendants normaux. On leur a fait écouter toutes les listes à un niveau de perception de 50 dB, dans le silence, d'une part, et dans le bruit, avec un rapport signal-à-bruit (S/B) de 10, 5, 0, -5, -10, -15 et -20 dB, d'autre part. On ne note aucune variation statistique entre les listes en ce qui concerne l'aptitude à reconnaître les mots lorsqu'on fait passer l'épreuve dans le silence (p = 0,542). Cependant, on note un effet de liste significatif lorsque l'épreuve est administrée dans le bruit (p > 0,0082). Généralement, de telles différences servent à appuyer l'avertissement selon lequel il convient de tempérer l'interprétation des variations de rendement d'un sujet à l'autre lors de l'exécution de tâches verbales avec un bruit de fond. Toutefois, en pratique, les écarts sont si faibles (p. ex., on ne peut expliquer que moins de 1 % de la variance; l'écart moyen entre les listes n'excède jamais 7,9 % ou est inférieur à quatre mots) qu'on considère l'effet sans importance d'un point de vue comportemental et/ou clinique.

Word recognition is routinely used by audiologists as part of a basic battery of tests to measure auditory function and for hearing aid evaluations. Numerous stimulus lists are available and there are several procedures for presenting test items and recording responses (see Bess, 1983; Olsen & Matkin, 1991). One such measure, which has enjoyed considerable usage in both clinical assessment (see Martin & Morris, 1989) and research, is the *Northwestern University Auditory Test No. 6 (NU-6)* originally reported by Tillman and Carhart (1966).

The stimuli contained in the *NU-6* are monosyllabic words having a consonant-nucleus-consonant construction. They are based on stimulus items originally developed by Lehiste and Peterson (1959). Initially, the *Northwestern University Auditory Test* comprised two lists (Tillman, Carhart, & Wilber, 1963) and was later expanded to four phonemic 50-word lists (Tillman & Carhart, 1966).

Open set word recognition testing in quiet has been criticized on a number of grounds including: a lack of face validity, an inability to differentiate normal and sensorineural impaired listeners, and an inability to differentiate the performance of hearing aids (Bess, 1983; Danhauer, Doyle, & Lucks, 1985; Surr & Schwartz, 1980). Assessment of word recognition ability in competing background noise has been advocated as a means of increasing face validity (Bess, 1983; Olsen & Matkin, 1991) and improving test sensitivity (Cohen & Keith, 1976; Cooper & Cutts, 1971; Findlay, 1976). The most common type of background competing stimulus is continuous broad band noise, due to its availability on clinical audiometers (Bess, 1983).

An important issue with respect to multiple list measures of word recognition abilities is the equivalence of test lists. Tillman and Carhart (1966) reported interlist equivalence for the *NU-6* stimuli with normal and hearing impaired listeners, and several authors have since verified these findings (Rintelmann, et al., 1974; Wilson, Coley, Haenel, & Browning, 1975). The reported list equivalency, however, has been reported only for data collected in quiet, and may not apply to situations where word recognition measures are employed in conditions with the presence of competing background noise.

It has been found that the presence of background noise can change the equivalency of speech materials (Chermak, Pederson, & Bendal, 1984; Chermak, Wagner, & Bendel, 1988; Gengel, Miller, & Rosenthal, 1981; Loven & Hawkins, 1983; Ripply, Dancer, & Pittenger, 1983; Schubert & Stenhjem, 1978). That is, the presence of noise has altered performance of speech materials to a greater extent than would be predicted for masking effects alone. With respect to phonemically balanced monosyllabic speech materials, it is speculated that competing noise changes a word list in a nonpredictive fashion and, consequently, the relationship between word lists (Loven & Hawkins, 1983). James, Bowsher, and Simpson (1991) suggest:

> The non-equivalence of isophonemic lists can be attributed to the fundamental inability of articulatory description to provide accurate information about the acoustical features of the word-forms and the occurrence of context effects. These contextual factors present in the particular combination of features plays a large part in the subjective perceptibility of the lists. The 'robustness' of the complete word-forms dictates the overall intelligibility of the word lists, and affects their relative difficulty under certain conditions .... Thus controlling for content, featurally or phonetically, between lists does not necessarily achieve a balance of difficulty (p.120).

To the best of our knowledge there has been no attempt to explore the equivalency of the *NU-6* lists when administered in the presence of background noise. The purpose of the present study was, therefore, to examine the effect of continuous broad band noise on the list equivalency of the *NU-6* stimuli.

Method

Subjects

Forty-eight young adults (M =23.7 years, SD =2.2; 15 males and 33 females) served as subjects. All subjects presented normal hearing sensitivity, defined as having pure tone thresholds at octave frequencies from 250 to 8000 Hz and speech reception thresholds (SRTs) of 20 dB HL (American National Standards Institute, 1989) or better. As well, all subjects presented with normal middle ear function (American Speech-Language-Hearing Association, 1990).

Stimuli

The test stimuli consisted of custom two channel stereo cassette tape recordings of 50-word lists one to four of the *NU-6* (male talker). The competing stimulus was continuous broad band noise. The custom recordings were developed by first transferring compact disk male talker recordings of the *NU-6* lists (Department of Veterans Affairs, 1989) onto an IBM compatible computer's (Zenith Model Z-386/20) hard drive via a compact disk player (Sony Model 608ESD) interfaced with an analog input/output board (Dalanco Spry Model 250). The word lists were then edited to remove the carrier phrase and to reduce the interstimulus intervals from 4.2 to 3.0 s.

The competing continuous broad band noise was generated by the computer. An examination of a 10 s segment of the noise file, using signal processing software (Signal Technology Inc. Model Interactive Laboratory System V6.1), confirmed that the noise spectrum was "flat" within two dB from 100 to 8000 Hz. Two 1000 Hz calibration tones were then generated, one at full scale and the other at the normalized power level (i.e., 0 dB signal-to-noise level). The full scale calibration was utilized to ensure no overloading within the recording/playback chain. All speech and noise files were normalized to have equal power. As the software with the analog input/output board only permitted single channel recording and playback, the process of playback from the computer and recording onto tape was first accomplished with two video cassette recorders (Panasonic Models PVS 4960 and AG-1960). The first recorded the noise file while the second combined the noise and speech files. The final VHS tape was employed to produce cassette copies for experimental use.

All editing, noise generation, signal power measurement and normalization, calibration tone generation, and playback from the hard disk was accomplished with custom software. The sampling rate for all these computer based operations was 20000 Hz giving an effective bandwidth of 8000 Hz.

Apparatus

A double wall sound-treated audiometric suite (Industrial Acoustics Corporation), meeting specifications for permissible ambient noise (American National Standards Institute, 1977), served as the test environment. The recorded stimuli were routed from a stereo cassette deck (AKAI Model GX-R66) to a clinical audiometer (Grason Stadler GSI 10 Model 1710-9700) and presented to the subjects through a Telephonics TDH-50P supraaural earphone housed in an M-51 cushion. All noise reduction and filter systems on the stereo cassette deck were in the off position during testing.

Procedure

Each subject was presented with the *NU-6* stimuli at 50 dB sensation level *re* their respective SRTs. The speech stimuli were presented in both quiet and noise with signal-to-noise ratios (S/Ns) of 10, 5, 0, -5, -10, -15, and -20 dB. The presentation order of lists, noise/quiet condition, and S/N was determined with a Latin Square design. All test stimuli were presented monaurally to the subjects' right ear. Subjects' responses were scored as total whole word percent correct by the test administrators.

Results

Total whole word percent correct mean scores and standard error of the means as a function of *NU-6* list and S/N are displayed in Figure 1. Total whole word percent correct mean scores and standard error of the means as a function of *NU-6* list in quiet are shown in Table 1. Not surprisingly, subjects' performance improved with increasing S/Ns and the best performance was observed in the quiet condition.

Figure 1. Total whole word percent correct mean scores in continuous broad band noise as a function of S/N and *NU-6* list. Error bars represent plus/minus one standard error of the mean.

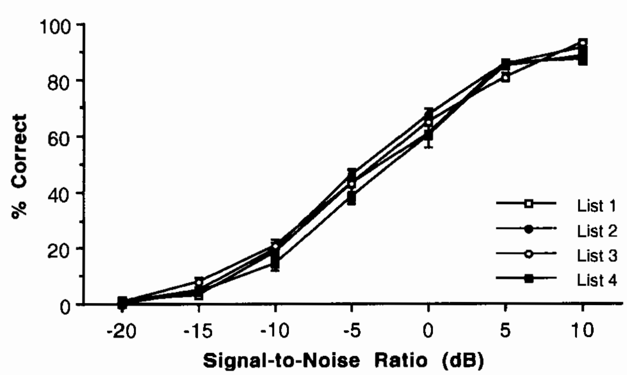

Table 1. Total whole word percent correct mean scores and standard errors of the mean (SEM) in quiet as a function of *NU-6* list.

List	*M*	*SEM*
1	93.5	1.3
2	94.7	0.9
3	92.7	0.8
4	94.7	0.9

The subjects' proportional scores were transformed to rationalized arcsine units [i.e., a simple linear arcsine transformation (Studebaker, 1985)] prior to subjecting them to inferential statistical analyses. In order to investigate the effect of *NU-6* list and S/N on word recognition performance, a two-way mixed analysis of variance (ANOVA) was conducted. List was the between subjects factor while S/N was the within subjects factor. Significant main effects of List [$F(3, 44) = 4.45$, $p = .0082$] and S/N [$F(6, 264) = 1072.83$, $p < .0001$] were found as well as a nonsignificant interaction between the two factors [$F(18, 264) = .45$, $p = .45$]. Percentage of variance accounted for each source was .16, 94.8, and .22 for List, S/N, and List X S/N respectively. A separate one-way ANOVA was also utilized to determine the effect of *NU-6* list on word recognition performance in quiet. The results indicated no statistically significant differences in word recognition as a function of list [$F(3,44) = .73$, $p = .542$].

As our primary interest was to investigate *NU-6* list effects, we further examined the main effect of list with Student-Newman-Keuls post hoc pair-wise comparisons between lists at each S/N. Statistically significant differences between pairs of list means were found at S/Ns of 10, 5, -10, and -15 dB, while no statistically significant differences among pairs were found at S/Ns of 0, -5, and -20 dB. The results of these comparisons which found significant differences are displayed in Table 2.

Table 2. Statistically significant ($p < .05$) Student-Newman-Keuls pair-wise comparisons of *NU-6* list transformed means as a function of S/N.

List	1	2	3	4
1				
2				
3	a, d		b	
4		a, c		

Note: a, b, c, and d denote pair-wise comparisons at S/Ns of 10, 5, -10, and -15 respectively.

It is worthy to note that although a list effect was found to be statistically significant, the effect size was very small indeed. That is, the amount of data variance accounted for by these differences was less than 1%. Further, across all S/Ns the greatest difference between list means was 7.9% (i.e., less than four words). On average, mean list differences across S/Ns were 5.4% or less than three words. Moreover, the directions of the differences appear not to be systematic.

Discussion

The findings of the present study suggest that, although interlist equivalency of *NU-6* lists one to four exists in quiet, the lists are not necessarily equivalent when administered in continuous broad band noise. Findings of nonequivalence of word lists when administered in noise is concordant with other investigators who report nonequivalence of other word lists administered in competing background noise (Chermak et al., 1984; Gengel et al., 1981; Loven & Hawkins, 1983). The findings *NU-6* list equivalency in quiet are consistent with previous reports (Rintelmann et al., 1974; Tillman & Carhart, 1966; Wilson et al., 1975).

We are of the opinion that the list differences seen here, while statistically significant, are behaviorally/clinically irrelevant. This interpretation is based on three lines of evidence. First, the size of the list effect was small, accounting for less than 1% of the data variance. Mean list differences across S/Ns never exceeded 7.9% (i.e., mean list differences were less than four words). Second, the direction of mean list differences appeared to be random. Finally, it has been suggested that when a between-subjects designed test is employed to investigate list equivalency, test variability can not be estimated (Dillon, 1982). That is, "when scores from different subjects are combined and treated as repeated measures of the same subject, the dispersion of scores so obtained cannot be attributed to the 'variability' of the test" (Dillon, p. 55). We suggest, however, that this is only important for "equivalency" studies that actually find list differences. In other words, when differences are found one does not know if they result from list nonequivalences or from intersubject variability. The present study found very small and unsystematic list effects.

It should be cautioned that conclusions regarding the stimuli used in this study may not be appropriately applied to other *NU-6* applications. For example, it has been reported that the spectra of *NU-6* stimuli differ considerably depending on the talker and recording procedure (Sherbecoe, Studebaker, & Crawford, 1993). As well, stimuli vary in their presentation level because of the difference between the long-term RMS level of the list material and the calibration tone (Sherbecoe, et al.). Further, when one employs a different competing background noise, the relationship between the articulatory composition of the words and the spectral characteristics of the noise changes thereby influencing masking effects. As well, Schwartz, Bess, and Larson (1977) reported that the individual variability of the *NU-6* for half-list measures at various S/Ns is too large to support its use. Finally, clinicians should also be aware of the possibility that individuals with varying degrees of hearing impairment may not mirror the pattern of performance indicated by this group of normal listeners (Loven & Hawkins, 1983). Considering the above, it seems prudent to question the list equivalency of *NU-6* stimuli as a function of recorded stimuli, competing background noise, list length, and clinical population.

Please address all correspondence to: Andrew Stuart, Dept. of Psychology, Dalhousie University, Halifax, Nova Scotia B3H 4J1.

References

American National Standards Institute (1989). *Specifications for audiometers.* (ANSI S3.6-1989). New York: ANSI.

American National Standards Institute (1977). *Criteria for permissible ambient noise during audiometeric testing.* (ANSI S3.1-1977). New York: ANSI.

American Speech-Language-Hearing Association (1990). Guidelines for screening for hearing impairments and middle ear disorders. *Asha, 32 (Suppl. 2)*, 17-24.

Bess, F.H. (1983). Clinical assessment of speech recognition. In D.F. Konkle & W.F. Rintelmann (Eds.) *Principals of speech audiometry* (pp. 127-202). Baltimore, MD: University Park Press.

Chermak, G.D., Pederson, C.M., & Bendel, R.B. (1984). Equivalent forms and split-half reliability of the NU-CHIPS in noise. *Journal of Speech and Hearing Disorders, 49,* 196-201.

Chermak, G.D., Wagner, D.P., & Bendel, R.B. (1988). Interlist equivalence of the Word Intelligibility by Picture Identification Test administered in broad-band noise. *Audiology, 27,* 324-333.

Cohen, R.L., & Keith, R.W. (1976). Use of low-pass noise in word recognition. *Journal of Speech and Hearing Disorders, 19,* 48-54.

Cooper, J.C., & Cutts, B.P. (1971). Speech discrimination in noise. *Journal of Speech and Hearing Research, 14,* 332-337.

Danhauer, J.L., Dolye, P.C., & Luks, L. (1985). Effects of noise on NST and *NU-6* stimuli. *Ear and Hearing, 6,* 266-269.

Department of Veterans Affairs (1989). *Speech recognition and identification materials, disc 1* [Compact Disc]. Long Beach, CA.

Dillon, H. (1982). A quantitative examination of the source of speech discrimination test score variability. *Ear and Hearing, 3,* 51-58.

Findlay, R.C. (1976). Auditory dysfunction accompanying noise-induced hearing loss. *Journal of Speech and Hearing Research, 14,* 374-380.

Gengel, R.W., Miller, L., & Rosenthal, E. (1981). Between and within listener variability to CID W-22 presented in noise. *Ear and Hearing, 2,* 78-81.

James, C.J., Bowsher, J.M., & Simpson, P.J. (1991). Speech audiometry: digitization effects and the non-equivalence of isophonemic word lists. *British Journal of Audiology, 25,* 111-121.

Lehiste, I., & Peterson, G.E. (1959). Linguistic considerations in the study of speech intelligibility. *Journal of the Acoustic Society of America, 31,* 280-286.

Loven, F.C., & Hawkins, D.B. (1983). Interlist equivalency of the CID W-22 word lists presented in quiet and in noise. *Ear and Hearing, 4,* 91-97.

Martin, F.N., & Morris, L.J. (1989). Current audiologic practices in the United States. *Hearing Journal, 42* (4), 25-44.

Olsen, W.O., & Matkin, N.D. (1991). Speech audiometry. In W.F. Rhintelmann (Ed.), *Hearing assessment* (2nd ed.) (pp. 99-140). Austin, TX: Pro-Ed.

Rintelmann, W.F., Schumaier, D.R., Jetty, A.J., Burchfield, S.A., Beasley, D.S., Mosher, N.A., Mosher, R.A., & Penley, E.D. (1974). Six experiments on speech discrimination utilizing CNC monosyllables. *Journal of Auditory Research, 2* (Suppl.), 1-30.

Ripply, J.V., Dancer, J.E., & Pittenger, J.B. (1983). List equivalency of the CID Everyday Sentences (Harris Revision) under three signal-to-noise ratios. *Ear and Hearing, 4,* 251-254.

Schubert, G.W., & Stenhjem, B.W. (1978). A reliability study of the Goldman-Fristoe-Woodcock test of auditory discrimination with learning disabled children. *Acta Symbolica, 7-9,* 43-56.

Schwartz, D.M., Bess, F.H., & Larson, V.D. (1977). Split half reliability of two word discrimination tests as a function of primary-to-secondary ratio. *Journal of Speech and Hearing Disorders, 42,* 440-445.

Sherbecoe, R.L., Studebaker, G.A., & Crawford, M.R. (1993). Speech spectra for six recorded monosyllabic word lists. *Ear and Hearing, 14,* 104-111.

Studebaker, G. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research, 28,* 455-462.

Surr, R.K., & Schwartz, D.M. (1980). Effects of multitalker competing speech on variability of the California Consonant Test. *Ear and Hearing, 1,* 319-323.

Tillman, T.W., & Carhart, R. (1966). *A test for speech discrimination composed of CNC monosyllabic words. Northwestern University Auditory Test No. 4* (Technical Documentary Report No. SAM-TDR-62-135). Brooks Air Force Base, TX: USAF School of Aerospace Medicine Technical Report.

Tillman, T.W., Carhart, R., & Wilber, L. (1963). *An expanded test for speech discrimination utilizing CNC monosyllabic words. Northwestern University Auditory Test No. 6.* Brooks Air Force Base, TX: USAF School of Aerospace Medicine Technical Report.

Wilson, R.H., Caley, K.E., Haenel, J.L., & Browning, K.M. (1975). Northwestern University Auditory Test No.6: Normative and comparative intelligibility functions. *Journal of the American Audiology Society, 1,* 221-228.

Author notes

A. Stewart is supported by the Medical Research Council of Canada and the Killam Trusts, Dalhousie University. D.P. Phillips is supported by grants from the Natural Sciences and Engineering Research Council of Canada. The authors gratefully acknowledge the helpful comments of two anonymous reviewers on an earlier draft of this manuscript.