A Frequency Importance Function for a New Monosyllabic Word Test

BELINDA A. HENRY, HUGH J. MCDERMOTT,
COLETTE M. MCKAY, CHRIS J. JAMES AND GRAEME M. CLARK
The University of Melbourne

A frequency importance function, characterising the relative contribution of different frequency bands to speech intelligibility, was determined for a CNC monosyllabic word test designed for Australian usage at the University of Melbourne. The importance function was derived from the phoneme scores of 12 normally-hearing listeners who were tested under various conditions of low- and high-pass filtering presented at signal-to-noise ratios of –8 to +6 dB, using noise which was shaped across frequency to match the speech spectrum. The importance function showed a dominant peak at approximately 2000 Hz, which is consistent with previously published word-test importance functions. The word test, along with the importance function, will be useful in advanced hearing-aid fitting procedures and research aimed at improving speech perception.

A new CNC (Consonant — vowel Nucleus — Consonant) monosyllabic word test has been developed at the University of Melbourne, and is used extensively in this laboratory for the assessment of hearing impaired subjects. The purpose of this report is to describe the long-term spectral characteristics, and the distribution of speech information across frequency, for this speech material.

The Speech Intelligibility Index (SII) is a measure of the amount of information contained in the speech signal that is available to a listener. The amount of speech information is dependent on two factors: how much of the speech spectrum is audible, and the importance of the audible portion for speech intelligibility. The SII is based on Articulation Index (AI) theory, originally proposed by French and Steinberg (1947). The SII is defined by the equation:

$$SII = \sum_{i} I_i W_i \qquad (1)$$

In this equation, I_i and W_i are the importance (I) and the audibility (W) associated with the frequency band i. The importance function (how I_i varies across frequency) represents the relative concentration of speech information in the different frequency bands (given that they are all audible). W_i is the proportion of the speech dynamic range which is audible within each band. The calculated SII, a value between 0.0 and 1.0, is the proportion of the total information present in the speech signal that is available to the listener. Speech intelligibility (as measured by a score on a speech test, for example) can be predicted from the SII. The relationship between SII and speech intelligibility is called the transfer function.

Correspondence and reprint requests: Belinda A. Henry, The University of Melbourne, Department of Otolaryngology, Parkville VIC 3052, Australia.

There has been renewed interest over recent years in the use of the SII to quantify the effect of the audibility of the speech signal on speech perception ability. Numerous studies have investigated the use of the SII as a tool to predict the ability of hearing-impaired listeners to perceive speech and, in particular, as a method for predicting performance with a hearing aid (Ching et al., 1997; Dirks et al., 1986; Dubno et al., 1989; Humes et al, 1986; Kamm et al., 1985; Ludvigsen, 1987; Pavlovic, 1984; Pavlovic et al., 1986; Schum et al., 1991; Skinner et al., 1982; Skinner & Miller, 1983).

In order to illustrate the use of the SII, we have constructed a figure (Fig. 1) which shows the aided thresholds of a hypothetical hearing-aid user, along with the long-term average spectrum of the words used in this study for an overall level of 70 dB SPL.

The audibility of the speech signal can be determined from this graph. However, using the extra information in the frequency importance function which will be derived in this paper, one can also predict which frequency region would contribute most to speech understanding for this listener (see results section), and which hearing aid frequency response would maximise the amount of speech information audible. This hypothetical listener will be used later to illustrate the results of the SII measurements.

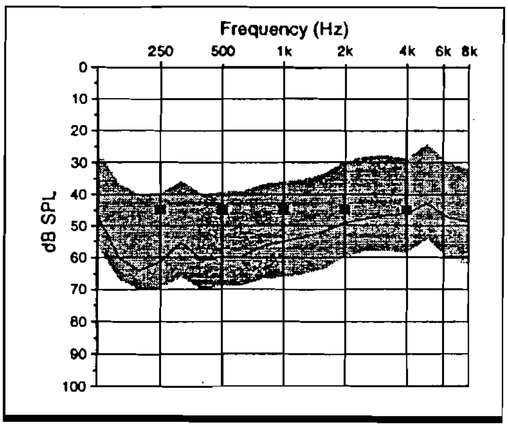

FIGURE 1
Aided thresholds of a hypothetical hearing-aid user (filled squares), along with the spectrum of the CNC words used in this study. The shaded region represents the nominal speech dynamic range of 30 dB with the highest level being the 1% speech-peaks curve. The solid curve is the long-term average spectrum. Both spectrum measurements used one-third octave bands.

One useful aspect of measuring and using SII instead of speech perception scores is that SII values for different frequency bands can be added or subtracted directly to estimate the SII for the combined bands. For example, if the speech perception score was measured for a person listening to speech limited to frequencies above and below 1 kHz respectively, her/his speech score for the whole signal would not be predictable by adding the two scores. In contrast, the SII value for the whole signal would be the simple sum of the SIIs for the two frequency regions calculated separately, and could be used (via the transfer function) to predict the overall speech recognition score.

The importance function, and the shape of the transfer function, are dependent on the speech material and the talker (Bell et al., 1992; DePaolis et al., 1996; Duggirala et al., 1988; Pavlovic, 1987; Studebaker et al., 1987; Studebaker & Sherbecoe, 1991; Studebaker et al., 1993), so should be derived individually for any specific speech test that is used in SII predictions. However, to predict speech intelligibility across a wide variety of listening situations (such as for general hearing-aid usage outside the laboratory), an importance function for 'average speech' is available (ANSI S3.5–1997).

In practice, the importance function must be derived experimentally using speech perception tests under various conditions of filtering and masking. The filtering involves several high-pass and low-pass conditions as well as an all-pass condition, and the masking varies the proportion of the speech dynamic range which is audible within each frequency band. These speech perception tests reduce the amount of information available from the signal in a controlled manner so that the importance of different frequency regions can be determined. In the following, we describe how the importance and transfer functions were determined for the new CNC word test.

METHOD

Subjects

Twelve normally-hearing adults (3 female, 9 male), all native Australian English speakers between 19 and 27 years of age, participated in the experiments. Normal hearing was defined as having pure-tone air conduction thresholds ≤ 15 dB HL (ANSI S3.6-1989) at octave frequencies from 125 to 8000 Hz in the ear used in the experiment.

Stimuli

The speech material was a CNC (Consonant — vowel Nucleus — Consonant) monosyllabic word test. There are 30 lists, each of 50 meaningful words, spoken by a female talker having an average Australian accent. Each list contains an identical phoneme set, which is based on that of Peterson and Lehiste (1962), with adaptations for Australian usage (e.g. some consonants rarely used in Australian English, such as the final consonant /r/, were removed). Each list contains a different set of words, and no word is used more than once within each of three sets of 10 lists. The words were recorded digitally onto compact disc.

The long-term average speech spectrum was determined in each one-third octave band. For each of the 1500 words, estimates of the spectrum were made using a 1024-point fast Fourier transform, with RMS levels obtained in one-third octave bands from 125 to 8000 Hz. The mean of these RMS levels was determined for each word, and these mean levels were averaged to determine the long-term average speech spectrum. The 1% speech peaks spectrum (the RMS levels exceeded by the speech signal 1% of the time) was determined in each one-third octave band, by ranking all of the RMS level measurements from all the words and then identifying the level which was exceeded 1% of the time. The long-term average speech spectrum and the 1% speech-peaks spectrum are shown in Fig. 1 as the solid line and higher-level limit of the shaded region, respectively.

In the experiment, the words were filtered, and presented in a continuous noise. There were 16 filter settings used: one wideband condition (80 – 10,240 Hz); seven high-pass (HP) conditions with low-frequency cut-offs of 482, 746, 1200, 1445, 1808, 2269, and 2859 Hz; and eight low-pass (LP) filter conditions, with seven using the same cut-off frequencies, plus an additional one with a 4566 Hz cut-off. Digital filters were used that provided an average frequency-response slope of 0.58 dB/Hz.

The masking noise was white noise which was shaped, using a custom-designed digital filter, to match the spectrum of the speech peaks. This noise masks an equal portion of the dynamic range of the speech signal at all frequencies. The noise was combined with the speech in a two-channel clinical audiometer (Madsen OB-822), which was used to set the levels and signal-to-noise (S/N) ratios of the stimuli. Eight S/N ratios were used with values between +6 and –8 dB in 2 dB steps.

The stimuli were delivered monaurally to each subject using an Etymotic Research insert earphone (ER-4B), which has a flat frequency response (± 2dB, relative to the sound field, between 50 Hz and 10 kHz). The long-term RMS level of the test words was 82 dB SPL, measured in an ear simulator (Bruel & Kjaer 4157), connected to a sound level meter (Bruel & Kjaer 2235). This level is equivalent to approximately 70–75 dB SPL in the free field.

Procedures

Prior to the experiment, the subjects were familiarised with the task. The subjects were randomly assigned to one of two groups consisting of 6 subjects each. Subjects in one group were tested at S/N ratios of +6, +2, –2, –6 dB, while subjects in the other group were tested at S/N ratios of +4, 0, –4, –8 dB. This was done to reduce the total test time for each subject. Each subject received one word list under each of the 16 filter conditions and at each of four S/N ratios (a total of 64 word lists). The order of presentation of both the word lists and the test conditions was determined by random selection without replacement. After exhausting the 30 lists, the selection procedure was repeated until the 64 conditions were completed. The subjects were instructed to repeat each word presented, and their responses were recorded. The percentage of phonemes correctly identified in each list was then calculated.

RESULTS AND DISCUSSION

The procedure used to determine the transfer and frequency importance functions from the raw data involved three steps, which closely followed those described by Studebaker and Sherbecoe (1991). First, the transfer function was derived in relative form, in which the highest-scoring condition (+6 dB S/N ratio and wideband filter) was assigned an SII value of 1.0, and the other conditions were assigned SII values relative to this. Secondly, the frequency importance function was derived by converting the speech perception scores for each frequency band into SII values using the relative transfer function. Thirdly, the absolute transfer function was derived by determining the appropriate SII for the highest-scoring condition.

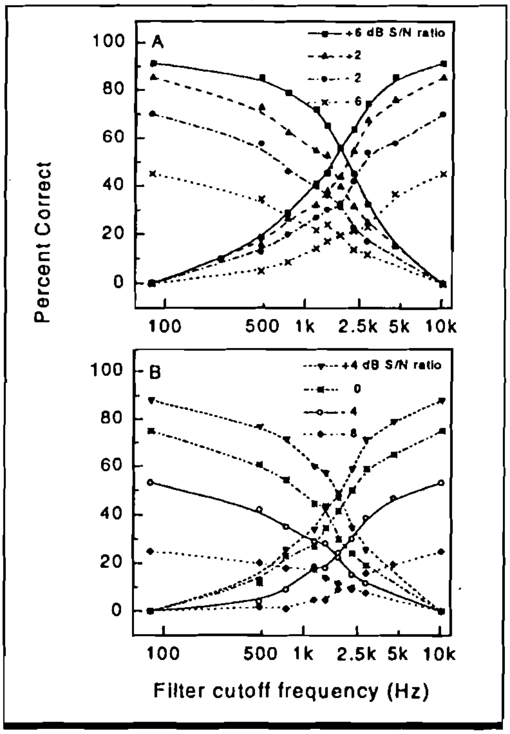

FIGURE 2
Mean scores (% phonemes correctly identified) versus filter cut-off frequency for high-pass and low-pass filter conditions. Smoothed curves were derived using a cubic B-spline. **(A)** S/N ratios +6, +2, −2, −6 dB; **(B)** S/N ratios +4, 0, −4, −8 dB.

Mean Scores

The phoneme scores were averaged across subjects for each filter condition and S/N ratio. Fig. 2 shows these mean scores, plotted for each S/N ratio as a function of filter cut-off frequency.

As shown in the figure, the data were smoothed using a cubic B-spline (a fitted cubic function which smoothes the data points using data on either side of each point). The two subject groups are shown separately for clarity. When the data from the two panels in Fig. 2 are plotted together, the data points for each particular S/N ratio from panel B (for example +4 dB) are evenly situated between the data points for the adjacent S/N ratios from panel A (for example +2 and +6 dB), indicating that the data from the two subject groups did not differ in any systematic subject-dependant way.

Derivation of the Relative Transfer Function

The smoothed mean score curves were used to derive the test score resulting from various SII values (assigning the highest-scoring condition an SII equal to 1.0). A double exponential function (the relative transfer function) was fitted to the test-score-versus-SII data. This function, which predicts the test-score from the SII, has the same form as the absolute transfer function (Eq. 2 below), except for different numerical constants. The fitting procedure determines the numerical constants in the function. Similarly, when the SII is plotted against test-score, a function can be fitted which predicts SII from the test-score (similar to Eq. 3 below).

Derivation of the Frequency Importance Function

The smoothed mean scores for each filter condition and S/N ratio were converted into SII values using the relative transfer function. The SII values obtained were used to determine the amount of speech information contained in each frequency region defined by the area between filter cut-off frequencies. This was done for each S/N ratio by averaging two estimates, one based on the HP data, and the other based on the LP data. For the HP data, the SII value for the higher cut-off frequency was subtracted from that for the lower cut-off frequency, while a corresponding procedure was applied to the LP data. The result in each case was an estimate of the amount of speech information (SII) in the region between the two cut-off frequencies. An average cumulative curve was determined by cumulating the mean SII values across frequency for each S/N ratio, and then averaging those curves. The average curve was then normalised to the range 0 to 1 by dividing each SII by the total cumulative SII. Finally, the relative amount of speech information in each one-third octave band (referred to as the weight) was determined by subtracting the interpolated SII value for the low-frequency edge of the band from that of the high-frequency edge of the band. These weights, which represent the proportion of the total SII contained in each one-third octave band, are shown in Table I.

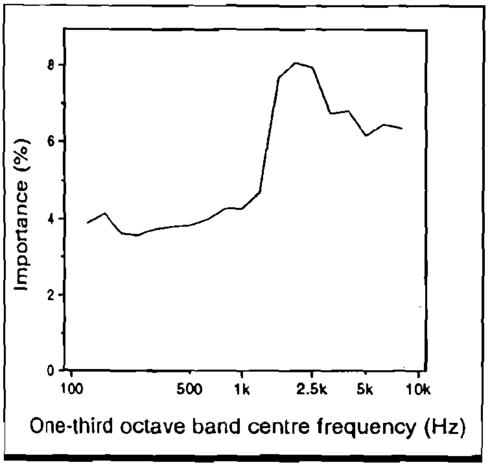

FIGURE 3
The frequency importance function (for one-third octave bands) for the CNC word test.

The frequency importance function, as a graph of these weights versus frequency, is shown in Fig. 3.

The importance function shows a peak in importance centred at approximately 2000 Hz. The primary importance to word intelligibility of frequencies around 2000 Hz has been demonstrated by several authors [CID W-22 test (Studebaker & Sherbecoe, 1991); NU-6 test (Studebaker et al., 1993); PB-50 test (DePaolis et al., 1996)]. This frequency

TABLE I
The Frequency Importance Function for CNC Words in One-third Octave Bands

1/3 octave band centre frequency (Hz)	Importance weight (SII × 100) (%)
125	3.90
160	4.15
200	3.63
250	3.57
315	3.74
400	3.80
500	3.83
630	4.00
800	4.27
1000	4.27
1250	4.67
1600	7.67
2000	8.07
2500	7.94
3150	6.74
4000	6.80
5000	6.15
6300	6.45
8000	6.35

Note. Importance weights are expressed as percentages.

region probably provides a greater contribution to speech intelligibility because it contains information about formant frequencies and their transitions, which are important for both vowel and consonant identification. Similarly, the cross-over frequency (at which half the information lies at higher or lower frequencies) for our word test was 1770 Hz, which is broadly consistent with that found for the other word tests listed above. Factors other than the words used, or the speaker, can contribute to differences in the frequency of maximum information or the cross-over frequency. These include the quality of the recording and the frequency response of the headphones used. It is interesting to note that there appears to be more information in the highest and lowest frequency regions for our data, compared to that of the above studies. This could be due to many factors including a combination of high-quality recording and headphones with a wide frequency response, and the speaker characteristics. Also, the shape of the importance function at the frequency extremes will depend on assumptions made in the analysis about how to interpolate the predicted score for regions between the full-band exclusion (zero correct) and the scores obtained for the narrowest pass bands used at either frequency extreme (see Fig. 2).

Returning to our example in Fig. 1, we can now compare the amount of speech information available to the listener at different frequencies. For the one-third octave centred at 2 kHz, the audibility is 0.5 (the threshold is at the mid-dynamic range point of the speech spectrum) and the frequency importance weight is 0.08 (Table I), giving an SII of 0.04 for this band. At 250 Hz, the audibility is higher (0.8) but the importance is lower (0.036), and therefore the SII is also lower (0.03). Thus the 2 kHz region provides more speech information to this listener than the 250 Hz region, in spite of the lower audibility at 2 kHz.

Derivation of the Absolute Transfer Function

The derivation of the absolute transfer function closely followed the procedure of Studebaker et al.,(1993). The procedure adjusted the relative transfer function by finding the appropriate value for the SII in the highest-scoring condition (which was set to 1.0 for the relative transfer function). In the wide-band filter condition, the SII was limited solely by the audibility, and hence equalled the proportion of the speech dynamic range (assumed to be 30 dB) which was audible above the masking. This proportion was estimated from the speech spectrum measurements, and then adjusted in an iterative fitting procedure. This fitting procedure found the value of SII for the wide-band condition which produced the best fit (minimum χ^2) of the SII versus test-score data. The value of SII determined for this condition was 0.68.

The resultant absolute transfer function for the prediction of test-score from SII is described below.

$$P = (1 - 10^{-SII/0.474})^{2.518} \qquad (2)$$

In this equation, P is the speech perception test-score expressed as a proportion, and SII is a value calculated from the audibility and importance functions (Eq. 1).

Similarly, the absolute transfer function for the calculation of SII from a measured test-score is:

$$SII = -0.445\log(1-P^{1/2.736}) \qquad (3)$$

These two functions are not exactly the inverse of each other, as might be expected. This is because, in the first case, the fitting procedure fitted the variance in the test-scores, while, in the second case, it fitted the variance in the SII values. It should be noted that these transfer functions have the same shape (but different numerical constants) as those previously found using whole-word scoring on word tests [CID W-22 test (Studebaker & Sherbecoe, 1991); NU-6 test (Studebaker et al., 1993); PB-50 test (DePaolis et al., 1996)].

Returning to the hypothetical listener represented in Fig. 1, we can now predict the test score he/she would achieve with the new CNC word test, by adding the SII contributions across all the audible frequencies and then using Eq. 2 to predict the score. Assuming no audibility outside the 250 to 4000 Hz range, the total SII is 0.42. Using Eq. 2, this predicts that the test score will be 71% phonemes correct. Using the same procedure, the test score can be predicted for alternative hearing-aid responses, and the results can be used to select the one which predicts the highest speech intelligibility, while maintaining acceptable overall loudness levels.

In practice, the SII predicts the maximum speech information audible to the listener. Hearing-impaired listeners with more than a mild to moderate sensorineural impairment have an impaired ability to make use of information present in the speech signal (due to factors such as distortion caused by high levels, impaired frequency selectivity, etc). Therefore predictions of speech perception ability using SII alone tend to overestimate the scores for these listeners. Similarly, cochlear implantees have more difficulty than normally-hearing listeners in making use of information present in the acoustic signal. This is due to the signal being processed and transformed to an electrical signal, as well as due to individual differences in such aspects as acoustic nerve survival patterns. Current research in this laboratory is utilising these word lists and the measured importance function to model the reduction in information transmission in different frequency regions for implantees compared to normally-hearing listeners.

CONCLUSIONS

The frequency importance function (Fig. 3) derived in this experiment for the new CNC monosyllabic word test has a peak in importance at approximately 2000 Hz. The frequency of this peak is close to that in importance functions derived for three previously published monosyllabic word tests. The absolute transfer function (Eqs. 2 and 3) and the importance weights (Table I) can be used in SII calculations for subjects tested with these word lists. These results will be useful in predicting the aided (or unaided) speech perception ability of hearing-impaired listeners for this particular CNC word test recording.

ACKNOWLEDGMENTS

This research was supported by a National Health and Medical Research Council project grant, "Speech perception by cochlear implantees: perceptual and related psychophysical studies", the Human Communication Research Centre, and the Cooperative Research Centre for Cochlear Implant, Speech and Hearing Research. The speech material was recorded by the Cooperative Research Centre Combionic Aid Program. The first author was supported by an Australian Postgraduate Award and a University of Melbourne Faculty of Medicine Scholarship. The authors wish to thank Justin Zakis for help in setting up the experiment, as well as all of the subjects who participated in this research.

REFERENCES

American National Standards Institute (ANSI). 1997. Methods for the calculation of the articulation index, ANSI S3.5 (American National Standards Institute, New York).

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

Bell, S.T., Dirks, D.D., & Trine, T.D. 1992. Frequency importance functions for words in high- and low-context sentences. *J. Speech Hear. Res.* 35, 950–959.

Ching, T., Dillon, H., & Byrne, D. 1997. Prediction of speech recognition from audibility and psychoacoustic abilities of hearing impaired listeners. In W. Jesteadt (Ed.) *Modeling Sensorineural Hearing Loss*. Mahwah, NJ, Erlbaum. 433–445.

DePaolis, R.A., Janota, C.P., & Frank, T. 1996. Frequency importance functions for words, sentences and continuous discourse. *J. Speech Hear. Res.* 39, 714–723.

Dirks, D.D., Bell, T.S., Rossman, R.N., & Kincaid, G.E. 1986. Articulation index predictions of contextually dependent words. *J. Acoust. Soc. Am.* 80, 82–92.

Dubno, J.R., Dirks, D.D., & Ellison, D.E. 1989. Stop-consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss. II: Articulation index predictions. *J. Acoust. Soc. Am.* 85, 355–364.

Duggirala, V., Studebaker, G.A., Pavlovic, C.V., & Sherbecoe, R.L. 1988. Frequency importance functions for a feature recognition test material. *J. Acoust. Soc. Am.* 83, 2372–2382.

French, N.R., & Steinberg, J.C. 1947. Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Am.* 19, 90–119.

Humes, L.E., Dirks, D.D., Bell, T.S., Ahlstrom, C., & Kincaid, G.E. 1986. Application of the articulation index and the speech transmission index to the recognition of speech by normal-hearing and hearing-impaired listeners. *J. Speech Hear. Res.* 29, 447–462.

Kamm, C.A., Dirks, D.D., Bell, S.S. 1985. Speech recognition and the Articulation Index for normal and hearing-impaired listeners. *J. Acoust. Soc. Am.* 77, 281–288.

Ludvigsen, C. 1987. Prediction of speech intelligibility for normal-hearing and cochlearly hearing-impaired listeners. *J. Acoust. Soc. Am.* 82, 1162–1171.

Pavlovic, C.V. 1984. Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment. *J. Acoust. Soc. Am.* 75, 1253–1258.

Pavlovic, C.V. 1987. Derivation of primary parameters and procedures for use in speech intelligibility predictions. *J. Acoust. Soc. Am.* 82, 413–422.

Pavlovic, C.V., Studebaker, G.A., & Sherbecoe, R.L. 1986. An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals. *J. Acoust. Soc. Am.* 80, 50–57.

Peterson, G.E., & Lehiste, I. 1962. Revised CNC lists for auditory tests. *J. Speech Hear. Dis.* 27, 62–70.

Schum, D.J., Matthews, L.J., & Lee, F. 1991. Actual and predicted word-recognition performance of elderly hearing-impaired listeners. *J. Speech Hear. Res.* 34, 636–642.

Skinner, M.W., Karstaedt, M.M., & Miller, J.D. 1982. Amplification bandwidth and speech intelligibility for two listeners with sensorineural hearing loss. *Audiology* 21, 251–268.

Skinner, M.W., & Miller, J.D. 1983. Amplification bandwidth and intelligibility for speech in quiet and noise for listeners with sensorineural hearing loss. *Audiology* 22, 253–280.

Studebaker, G.A., Pavlovic, C.V., & Sherbecoe, R.L. 1987. A frequency importance function for continuous discourse. *J. Acoust. Soc. Am.* 81, 1130–1138.

Studebaker, G.A. & Sherbecoe, R.L. 1991. Frequency-importance and transfer functions for recorded CID W-22 word lists. *J. Speech. Hear. Res.* 34, 427–438.

Studebaker, G.A., Sherbecoe, R.L., & Gilmore, C. 1993. Frequency-importance and transfer functions for the Auditec of St. Louis recordings of the NU-6 word test. *J. Speech Hear. Res.* 36, 799–807.