Use of the CID W22 as a South African English Speech Discrimination Test

Wayne J. Wilson and Selvarani Moodley

Department of Speech Pathology and Audiology
University of the Witwatersrand

ABSTRACT

South Africa currently lacks a pre-recorded South African English (SAE) specific speech discrimination test. In the absence of such a test, the SAE speaker recording (Tygerberg recording) of the American (USA) English (AE) CID W22 wordlists - in combination with the original American CID W22 normative data - is the most widely used alternative. The reliability and validity of this method, however, has never been formally assessed. This study assessed the performance of 15 normal hearing, female, first language SAE speakers on the first two full-lists of Tygerberg CID W22 recording at 20, 30, 40, 50, 60 and 70 dBSPL, and compared their scores to the American CID W22 wordlist normative data. Overall, the South African subjects performed worse than the original American normative data at the lower presentation intensities (< 50 dBSPL). Use of the Tygerberg CID W22 recording - with the original American CID W22 normative data - for near threshold assessment of SAE speaking subjects was therefore concluded to be problematic. Use at suprathreshold intensities (> 40 dBSPL), however, was considered a viable option. These results reiterate the need for large scale, South African specific normative studies for the CID W22 wordlists if they are to continue their role as the dominant speech discrimination wordlists in South Africa.

KEY WORDS: Special discrimination, CID W22 wordlists, American (USA) English, South African English.

INTRODUCTION

The CID auditory test W22, from the United States of America, represents the most widely used, basic audiological speech discrimination test in South Africa. The test's popularity appears to be due mainly to its widespread use in the USA only, however, as opposed to any formal research demonstrating the validity and reliability of its use in the South African context. These factors, support for and criticisms against the CID W22, and the general problems of applying a non-South African specific speech discrimination test to the South African population have been discussed previously (Wilson, Jones & Fridjhon, 1998).

The CID W22 was originally designed as an improvement on the Psycho-Acoustic Laboratories phonetically balanced 50 word lists (PAL-PB 50) speech discrimination test. It improved phonetic balance (Lehiste & Peterson, 1955), equality of difficulty between lists (Brewer & Resnick, 1983) and degree of familiarity (Brewer & Resnick, 1983; Hirsh et al., 1952). The test consists of 4 lists of 50 monosyllabic words with six randomizations of each list. The words were originally chosen on the basis of being representative of North American (USA) English, with all but one of the chosen words appearing on the Thorndike "List of Most Frequently Used Words" (Hirsh et al., 1952). The chosen words were also checked for phonetic balance according to studies of syllable-consonant-vowel distribution in American English (AE) (Hirsh et al., 1952).

Whilst originally created as 50 word full-lists, users of the CID W22 wordlists recognised almost immediately the benefits of reducing test time by administering 25 word half-lists (Martin & Forbis, 1978; Edgerton & Danhauer, 1979). These shorter lists remain in widespread clinical use despite being prone to higher variability (and therefore poorer reliability) and exacerbation of the already difficult phonetic balance problem (Ostergard, 1983).

The original CID W22 standardisation information was obtained on 15 normal hearing, AE speakers. All four lists were randomly presented at 10 dB intervals from 20 dBSPL to 70 dBSPL. The resultant group "performance-intensity" plot is shown in Hirsh et al. (1952) and represents the averaged performance of the subjects on all four full-lists at each presentation intensity. The actual scores obtained are not given and are extrapolated from the Hirsh et al. (1952) plot.

The success of the CID W22 wordlists in the USA lead to their introduction and widespread use in South Africa. In an attempt to make these standardised CID W22 wordlists more suitable for the South African English (SAE) speaking population, a SAE speaker recording of the original wordlists was made at Tygerberg Hospital in the Western Cape.

Technical information on the Tygerberg CID W22 recording is not available. What is known is that the recording is comprised of the 4 original American wordlists of 50 words each, all read by a South African adult male who is a first language speaker of SAE. Each word is preceded by the carrier phrase "say the word", the inter-stimulus interval

is 4 seconds and serves as the subject's response period.

It must be noted at this stage that a popular alternative to the Tygerberg CID W22 recording is to present the wordlists via monitored live voice (MLV). Whilst the MLV presentation may provide a partial solution to speaker-listener accent mismatch, it has been widely criticised for its poor test-retest reliability (Carhart, 1965; Brandy, 1966; Kreul, Bell & Nixon, 1969 and Northern & Hattler, 1974) and the inherent problem of comparing results using monitored live voice to standardisation information obtained using recorded versions of the test (Mendel & Danhauer, 1997). The recorded version of a test keeps the tester voice constant therefore improving test-retest reliability and allows valid comparisons to standardised normative data. Because of this, testing using the Tygerberg recording was preferred in this study.

South African normative data for the Tygerberg recording of the CID W22 has not been published. As a result, the performance of SAE speaking subjects on the Tygerberg recording, whether it be half-list of full-list performance, is typically compared to the original AE subject, averaged full-list normative data for the original American CID W22 recording. Such comparisons add obvious reliability and validity problems to the already stretched reliability and validity of using an American designed and standardised test on South African subjects.

Formal assessment of the reliability and validity of using the Tygerberg CID W22 wordlist recording - with the original American CID W22 normative data - to assess first language SAE speakers is needed if this protocol is to remain the dominant speech discrimination test protocol in South Africa.

METHODOLOGY

AIMS

This study aimed to:

1. Measure the performance of 15 female, first language South African English (SAE) speakers on the first two full lists, and the resultant first four half-lists, of the SAE recorded version (Tygerberg recording) of the CID auditory test W22 at stimulus intensities of 20, 30, 40, 50, 60 and 70dBSPL.
2. Compare the SAE speaker results to the original American CID W22 wordlist normative data.
3. Compare the performance of the South African subjects only, between full-lists and between half-lists at each presentation intensity, and within full-lists and half-lists between each adjacent presentation intensity.
4. Comment on the suitability of using the Tygerberg recording of the CID W22 wordlists, with comparison with the original American CID W22 normative data, to assess first language SAE speakers.

SUBJECTS

Subject selection criteria

Subjects were selected using a convenience sampling technique. For acceptance into the study, subjects were required to conform to the following criteria (as confirmed by audiometric testing and self-report): (i) Aged 17 - 40 years. This controlled for the well-recognised effects of paediatric and geriatric age ranges on performance on speech discrimination tasks (Hall, 1983). (ii) Female gender. This criterion was included for ease of subject selection, as it was deemed unlikely that a sample balanced for gender could easily be obtained. (iii) English as first-language. (iv) Resident in Gauteng, South Africa as according to Fuller (1987), subjects for research in the area of speech audiometry should be native to the local area. (v) Hearing thresholds ≤ 5dBHL at 250, 500, 1000, 2000, 4000 and 8000 Hz in the test ear. (v) No significant history, past or present, of: hearing impairment; speech or language impairment; tinnitus; ear infections; noise exposure; or; family history of hearing problems. (vi) No previous knowledge of, or experience regarding the CID W22 wordlists.

Subject description

An initial sample of 18 female, first language SAE speaking subjects, all resident in the Gauteng area, was obtained. Three subjects were excluded on the basis of pure tone threshold criteria. The final 15 subjects ranged in age from 18.6 to 31.4 years, with a mean age of 23.9 ± 4.4 years. Educational levels included two subjects with secondary school education and 13 subjects with or receiving tertiary level education. The better ear only was tested in each subject with a final 8 right and 7 left ears tested.

TEST ENVIRONMENT, EQUIPMENT AND RECORDED MATERIAL

Testing was conducted in a two-room soundproof booth at the University of the Witwatersrand Speech and Hearing Clinic in Gauteng. A one-way mirror enabled observation of subjects during testing. The Tygerberg CID W22 wordlists were presented via an Aiwa audiocassette player coupled with a Grason Stadler GSI16 audiometer. Presentation was through TDH39 headphones with MXAR41 cushions.

MEASUREMENT PROCEDURES AND DATA COLLECTION

Subjects first filled in a case history questionnaire. Pure tone air conduction thresholds for 250-8000 Hz were then obtained for both ears using a standard Hughson-Westlake test procedure. If all selection criteria were met, the subject's better ear was chosen for speech testing.

Speech discrimination testing was conducted using the first two of the four Tygerberg CID W22 pre-recorded full wordlists at 20, 30, 40, 50, 60 and 70 dBSPL. Administration of all four full-lists at each intensity was not conducted as it was thought that this would lead to unacceptably high levels of test duration and subject fatigue. All testing was conducted by a final year BA (Speech and Hearing Therapy) student who has had clinical training in audiometric procedures, under the supervision of an audiologist registered with Health Professions Council of South Africa.

The speech discrimination testing followed a similar protocol to that used by the Central Institute for the Deaf to normalize the original CID W22 wordlists (Hirsh et al., 1952). The two full-lists were split and presented as four half-lists at each of the six presentation intensities in a pseudo-randomised order to prevent a particular list from occurring several times in succession. The same set of instructions was given to all subjects through the headphones: "You are going to hear sentences. I want you to repeat the last word of each sentence."

Subjects' responses were recorded and scored on-line by the tester. An all-or-none scoring procedure was used where the subject had to perceive the entire word correctly in order to receive credit. A correct response obtained a score of 2% for the full lists and 4% for the half-lists.

Subjects were allowed to rest at their request at any time during the test procedure in order to reduce fatigue (Mendel & Danhauer, 1997).

DATA ANALYSIS

South African English speakers' scores for the two full and four half-list Tygerberg CID W22 wordlists were visually compared to the original American, averaged scores on the American CID W22 full-lists [as extrapolated from the graphs provided in Hirsh et al. (1952)]. Descriptive differences were identified.

The South African subjects' scores on the Tygerberg CID W22 wordlists were then analysed for differences between:

- Mean scores on the two full-lists, for each presentation intensity separately, using two-tailed t-tests for dependent samples at the 5% significance level.
- Mean scores on the four half-lists, for each presentation intensity separately, using one-way repeated measure ANOVA and Tukeys Honest Significant Difference analyses at the 5% significance level.
- Mean scores between each adjacent pair of presentation intensities (20-30, 30-40, 40-50, 50-60 and 60-70 dBSPL), for each full-list and half-list separately, using repeated Wilcoxon Matched Pairs analyses at the 1% significance level.
- Score variances between presentation intensities for each full and half-list separately, and between full and half-

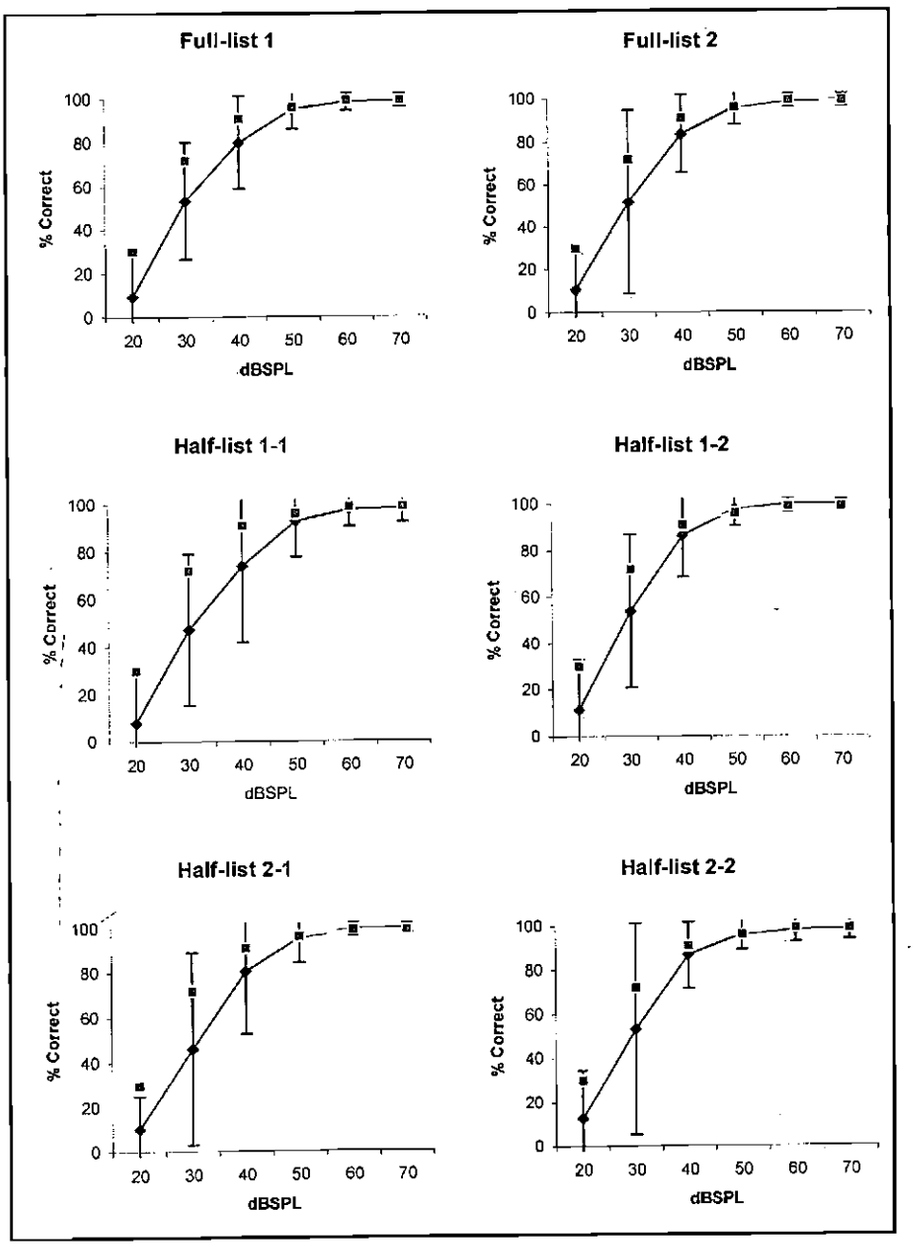

FIGURE 1. Plots of South African English speakers' mean scores (diamonds) with error bars (2 S.D) for each Tygerberg recording CID W-22 full-list and half-list, and the original American English speakers' mean scores (squares) for the American CID W-22 full-list.

lists for each presentation intensity separately, using within-groups F test analyses at the 1% significance level.

RESULTS

SOUTH AFRICAN ENGLISH SPEAKERS' VERSUS AMERICAN ENGLISH SPEAKERS' SCORES

Figure 1 displays a plot of the SAE speakers' mean scores on the Tygerberg CID W22 full-lists 1 and 2, and their respective half-lists 1-1, 1-2, 2-1 and 2-2, at the 20, 30, 40, 50, 60 and 70 dBSPL presentation intensities. All plots are superimposed on the extrapolated normative data for the AE speaker mean scores averaged over the four American CID W22 full-lists (Hirsh et al., 1952). In general, the SAE speakers' performance on the Tygerberg CID W22 wordlists was poorer than the AE speakers performance on the American CID W22 wordlists at the lower presentation intensities (< 50 dBSPL), but was equivalent at the higher presentation intensities (> 40 dBSPL).

On closer visual inspection, the SAE speakers' scores were:

- More than two standard deviations lower than the AE speakers' mean scores for half-list 2-1 at 20 dBSPL.
- Approximately two standard deviations lower for full-lists 1 and 2, and half-lists 1-1 and 1-2 at 20 dBSPL.
- Between one and two standard deviations lower for half-list 2-2 at 20 dBSPL and for full-list 1, and half-lists 1-1, 1-2 and 2-1 at 30 dBSPL.
- Less than one standard deviation lower for full-list 2 and half-list 2-2 at 30 dBSPL and all lists at 40 dBSPL.
- Approximately equal for all lists at 50, 60 and 70 dBSPL.

SOUTH AFRICAN ENGLISH SPEAKERS' SCORES

Table 1 shows the mean ± one standard deviation scores for 15 female, first language SAE speakers on the Tygerberg CID W22 full-lists 1 and 2, and their respective half-lists 1-1, 1-2, 2-1 and 2-2, at the 20, 30, 40, 50, 60 and 70 dBSPL presentation intensities.

The t-test for dependent sample results for differences between the SAE speaking subjects' full-list scores, conducted separately for each presentation intensity, showed no significant differences ($p<0.05$).

Table 2 shows one-way repeated measures ANOVA and Tukeys Honest Significant Difference (THSD) test results for differences between the SAE speaking subjects half-list scores, conducted separately for each presentation intensity. Results showed a significant difference ($p<0.05$) between the first half of list 1 (1-1), and the second half of list 1 (1-2) and list 2 (2-2) at 40dBSPL only. Note that a significant ($p<0.05$) ANOVA result was observed at the 60 dBSPL presentation intensity, but the associated THSD analysis showed no significant differences.

Wilcoxon Matched Pairs test results for differences between SAE speaking subject scores on each pair of adjacent presentation intensities (20-30, 30-40, 40-50, 50-60 and 60-70 dBSPL), conducted for each full and half-list separately, showed significant differences ($p<0.01$) for all comparisons, except between 60-70 dBSPL for all full and half-lists.

F test analysis for differences in variance between SAE speaking subjects scores, between presentation intensities (dBSPL) within each full and half-list, showed the following significant differences ($p<0.01$):

- (20, 30, 40) > 50 > (60, 70) for full-list 1.
- 30 > (20, 40) > 50 > (60, 70) for full-list 2.
- (20, 30, 40, 50) > (60, 70); and (30, 40) > 50 for half-list 1-1.
- (20, 30, 40) > 50 > (60, 70) for half-list 1-2.
- (20, 30, 40, 50) > (60, 70); and 30 > (20, 50) for half-list 2-1.
- 30 > (20, 40) > (50, 60, 70) for half-list 2-2.

F test analysis for differences in variance between SAE speaking subjects scores, between each full and half-list, within each presentation intensity, showed the following significant differences ($p<0.01$):

- half-list 1-1 > half-list 2-2 at 40 dBSPL.
- half-list 1-1 > (half-list 1-2, half-list 2-2) at 50 dBSPL.
- half-list 1-1 > (half-list 1-2, half-list 2-1) at 60 dBSPL.
- (half-list 1-1, half-list 2-2) > (half-list 1-2, half-list 2-1) at 70 dBSPL.

Whilst these differences have implications for the ANOVA and THSD analyses of the half-list scores above, the differences were not considered to be in excess of those tolerable by the ANOVA (Lindman, 1974).

DISCUSSION

The qualitative finding that the SAE speakers' scores on the Tygerberg CID W22 wordlists were generally worse than those of the original AE speakers' scores on the

TABLE 1: South African English speakers' mean ± one S.D. speech discrimination scores on the Tygerberg CID W22 full-lists 1 and 2, and their respective half-lists 1-1, 1-2, 2-1 and 2-2.

dBSPL	20	30	40	50	60	70
Full-list 1	9.2 ± 10.8	53.3 ± 13.8	80.1 ± 10.8	95.2 ± 4.7	98.5 ± 2.2	99.1 ± 1.8
Full-list 2	10.7 ± 8.5	51.6 ± 21.7	83.6 ± 9.3	95.9 ± 4.0	98.7 ± 1.6	99.2 ± 1.7
Half-list 1-1	7.7 ± 12.1	47.2 ± 15.9	73.9 ± 15.9	92.8 ± 7.4	97.6 ± 3.6	98.4 ± 2.9
Half-list 1-2	11.2 ± 10.7	53.9 ± 16.5	86.4 ± 9.3	97.6 ± 3.6	99.5 ± 1.4	99.7 ± 1.0
Half-list 2-1	10.1 ± 8.0	45.9 ± 21.7	80.5 ± 14.3	95.5 ± 5.4	99.5 ± 1.4	99.7 ± 1.0
Half-list 2-2	12.5 ± 10.9	53.1 ± 23.9	86.7 ± 7.5	96.0 ± 3.7	97.9 ± 2.6	98.7 ± 2.5

American CID W22 wordlists, indicates that the use of the Tygerberg SAE recording of the CID W22 wordlists - with the original American CID W22 wordlist normative data - cannot be applied to the SAE speaking population without significant modification.

Specifically, the differences between the SAE and AE speaking subjects on their respective CID W22 wordlists occurred at presentation intensities below 50 dBSPL. This suggests that both groups were able to hear their respective CID W22 wordlists equally well at the higher presentation intensities, but the American subjects were able to extract more information from their American wordlists at the lower presentation intensities, due most probably to their greater familiarity with AE. These findings are consistent with the previous finding of SAE speakers' performance on the Australian English NAL-AB wordlists (Wilson, Jones & Fridjhon, 1998).

Analysis of the SAE speakers' scores alone showed several useful findings. The finding that the mean scores showed only two significant differences ($p<0.05$) when compared within presentation intensities between lists (half-list 1-1 scores were lower than those of half-lists 1-2 and 2-2 at the 40 dBSPL presentation intensity only) suggests two things. Generally, it suggests that the two Tygerberg CID W22 wordlists and their four half-lists were of equal difficulty for the SAE speaking subjects at all but one of the tested presentation intensities. Specifically, it suggests that half-list 1-1 may be a more difficult half-list at 40 dBSPL. The possibility that half-list 1-1 could be a problematic list was further supported by the finding that this half-list had a significantly greater variance ($p<0.01$) than the other half-lists at the higher presentation intensities (> 30 dBSPL). Such interpretations must be approached with caution, however, as whilst the absolute differences in the mean and variance values were statistically significant, their clinical impact is of somewhat less value due to the small size of the absolute differences observed.

The SAE speakers' scores also showed a significant improvement ($p<0.05$) with each increase in presentation intensity from 20 to 60 dBSPL, within each Tygerberg CID W22 full and half-list, as would be generally expected in a normal performance intensity function. The absence of significant differences between the 60 and 70 dBSPL presentation intensities provides evidence of the beginning of a ceiling effect at these intensities.

TABLE 2: ANOVA and Tukeys HSD results for SAE speakers' speech discrimination scores on e Tygerberg CID W22 wordlists for half-lists 1-1, 1-2, 2-1 and 2-2.

dBSPL	ANOVA	THSD
20	p=0.38	–
30	p=0.18	–
40	p<0.001	1-1<(1-2,2-2)
50	p=0.06	–
60	p<0.05	–
70	p=0.06	–

The significant decrease ($p<0.05$) in the SAE speakers' score variances with increases in presentation intensity is remarkably similar to previous findings of SAE speakers' performance on the Australian NAL-AB wordlists (Wilson, Jones and Fridjhon, 1998). This finding, again, suggests a more equal ability between subjects to overcome any problems with non-SAE English, in this case AE, when given a high enough presentation intensity, and only partly agrees with Ostergard's (1983) findings that variability in speech discrimination scores tend to decrease for extreme scores and increase for mid-range scores.

Minor limitations to this study include the young, adult, female composition of the sample and the lack of control over right versus left ear selection. With no reports, to the authors' knowledge, of significant female/male or right/left ear performance differences in the literature for any of the monosyllabic word tests, these points were not considered to have had any significant impact on this study's findings.

A more significant limitation was prerequisite that the sample should be first language SAE speakers, and the predominance of tertiary level education. Davis (1983) (cited in Lutman, 1987) demonstrated a relationship between socio-economic status and type of occupation and performance in speech audiometry, with higher education level implying that a certain level of linguistic competence and even sophistication affects the results. It might therefore be predicted that a more representative sample of the SAE speaking population would not have performed as well as the more educated sample used in this study.

Finally, it must be noted a significant factor affecting any interpretation of speech discrimination results is the large amount of variability inherent in speech audiometric testing. Mendel and Danhauer (1997) warn that a margin of 16-20% should be allowed for erroneous scoring alone, because of errors in the scorer's perception. Similarly, Thornton and Raffin (1978), Ostergard (1983) and Green (1987) state that a single score obtained for a particular wordlist is only an indicator of a range of scores in which the true score is likely to lie. This inherent variability in speech audiometry diminishes its accuracy in all uses generally and the strength of the descriptive results of this study specifically.

CONCLUSIONS

The similarities observed between the SAE and AE speakers on their respective pre-recorded CID W22 wordlist recordings makes the current South African practice of using the Tygerberg SAE recording - with the original American normative data - a valid option for the speech discrimination assessment of SAE speakers, under restricted conditions. The Tygerberg CID W22 recording - with the American normative data - is most suitable for use at suprathreshold intensities (> 40 dBSPL) where the fewest differences were observed. Use at or near threshold (< 50 dBSPL) however, where the most differences were observed, should be approached with caution and reliance on these wordlists for site of lesion purposes should be avoided. In view of the generally equivalent performance of the SAE speakers within the Tygerberg CID W22 full and half-lists (a high "test equivalency"), this recording's general test reliability and validity could be quickly and easily improved by replacing the American normative data with large South African normative data-bases.

Despite this study's findings, the over-riding need for a

SAE specific speech discrimination test remains. Whilst the Tygerberg CID W22 wordlists - with the American normative data - was shown to have some validity and reliability in SAE speakers, significant limitations exist in its use, and the continued use of non-South African tests must be seen as an interim measure only.

In view of similarities between the performance of SAE and AE speakers on their respective CID W22 wordlists in this study, and similar findings in other wordlists (WIPI and NU-CHIPS (Mehl, 1992) and NAL-AB wordlists (Wilson, Jones and Fridjhon, 1998)), there is now a small but growing body of literature that can provide some of the groundwork needed to assist in the development of the long awaited SAE specific speech discrimination test.

REFERENCE LIST

Brandy, W.T. (1966). Reliability of voice tests of speech discrimination. *Journal of Speech and Hearing Research*, 9: 461-465.

Brewer, C.C. & Resnick, D.M. (1983). A review of tests of speech discrimination. *Seminars in Hearing*, 4: 205-220.

Carhart, R. (1965). *Problems in the measurement of speech discrimination. Archives of Otolaryngology*, 82(9): 253-260.

Edgerton, B.J. & Danhauer, J.L. (1979). *Clinical Implications of Speech Discrimination Testing Using Nonsense Stimuli.* Baltimore: University Park Press.

Fuller, H. (1987). Equipment for speech audiometry and its calibration. In M. Martin (Ed.), *Speech Audiometry*. London: Taylor and Francis Ltd.

Green, R. (1987). The uses and misuses of speech audiometry in rehabilitation. In M. Martin (Ed.), *Speech Audiometry*. London: Taylor and Francis Ltd.

Hall, J.W. (1983). Diagnostic applications of speech audiometry. *Seminars in Hearing*, 4: 179-204.

Hirsh, I.J., Hallowell, D., Silverman, S.R., Reynolds, E.G., Eldert, E. & Benson, R.W. (1952). Development of materials for speech audiometry. *Journal of Speech and Hearing Disorders*, 15:321-337.

Kreul, E.J., Bell, D.W. & Nixon J.C. (1969). Factors affecting speech discrimination test difficulty. *Journal of Speech and Hearing Research*, 12(2): 281-287.

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

Lindman, H.R. (1974). *Analysis of variance in complex experimental design.* San Fransisco: W.H. Freeman.

Lutman, M.E. (1987). Speech tests in quiet and noise as a measure of auditory processing. In M. Martin (Ed.) *Speech audiometry* (Chapter 3). London: Taylor & Francis Ltd.

Martin, F.N & Forbis N.K. (1978). The present status of audiometric practice: A follow up study. *ASHA*, 20: 531-541.

Mendel, L.L. & Danhauer, J.L. (1997). *Audiological evaluation and management and speech perception assessment.* San Diego: Singular Publishing Group Inc.

Northern, J.L. & Hattler, K.W. (1974). Evaluation of four speech discrimination test procedures on hearing impaired patients. *Journal of Aud*itory Research, (Suppl): 1-37.

Ostergard, C.A. (1983). Factors influencing the validity and reliability of speech audiometry *Seminars in Hearing*, 4: 221-240.

Thorton, A.R. & Raffin, M.J. (1978). Speech discrimination scores modelled as a binomial variable. *Journal of Speech and Hearing Research*, 21(3): 507-518.

Wilson, W., Jones, B. & Fridhon, P. (1998). Use of the NAL-AB wordlists as a South African English specific speech discrimination test. *South African Journal of Communication Disorders*, 45: 77-86.