

Use of the NAL-AB Wordlists as a South African English Speech Discrimination Test

Wayne J. Wilson, Beverley Jones and Peter Fridjhon

Department of Speech Pathology and Audiology,
University of the Witwatersrand
Department of Statistics and Actuarial Science,
University of the Witwatersrand

ABSTRACT

South Africa still lacks a South African English specific speech discrimination test. As an alternative, this study investigated the use of the Australian English, National Acoustic Laboratories Arthur Boothroyd (NAL-AB) wordlists to assess the speech discrimination of South African English speakers. Thirty South African English speakers were tested at 0, 5, 10, 20 and 25 dBHL (audiometer dial reading) and their performance-intensity functions were compared qualitatively to the NAL-AB wordlist normative data. Results showed three general patterns; similar performance for both groups; poorer performance by the South African English speakers at the low to mid presentation intensities only; and poorer performance by the South African English speakers across most presentation intensities. Use of the NAL-AB wordlists at threshold levels or for site of lesion assessment was therefore concluded to be unwise. Use of these wordlists at supra-threshold levels, however, would provide a valid and reliable option for the speech discrimination assessment of South African English speakers.

OPSOMMING

Daar bestaan tans nie 'n Engelse spraakdiskriminasie-toets spesifiek vir die Suid-Afrikaanse konteks nie. Hierdie studie was 'n ondersoek na die bruikbaarheid van die "National Acoustic Laboratories Arthur Boothroyd (NAL-AB)"-woordelyste, wat op Engelssprekende Australiërs toegepas word, om die spraakdiskriminasie van Engelssprekende Suid-Afrikaners te toets. Dertig proefpersone is by 0, 5, 10, 20, en 25 dBHL (oudiometerlesing) getoets, en hulle diskriminasie-intensiteitsfunksies is kwalitatief met die normatiewe data van die NAL-AB vergelyk. Daar is drie algemene patrone gevind: geen verskil tussen die twee groepe nie; swakker diskriminasie deur die Suid-Afrikaanse groep by slegs die lae en middel-intensiteite; en swakker diskriminasie deur die Suid-Afrikaanse Engelssprekendes by die meeste intensiteite. Die NAL-AB-woordelyste is egter geskik om die spraakdiskriminasie van Engelssprekende Suid-Afrikaners bo gehoordrempels te toets.

KEYWORDS: Speech discrimination, phonemic scoring, NAL-AB wordlists, Australian English, South African English.

South Africa still lacks a pre-recorded speech discrimination test specific to South African English. With no less than 11 current, official languages, and an unknown number of dialectal variations, the development of such a test that is both language and culturally appropriate is a formidable task. Of further hindrance is the need for community based definitions of "normal" language and communication (Pakendorf & Alant, 1997), and for ethnographically based research that defines current "South African English" language and culture (Taylor, 1986). Solving these prerequisite problems is proving to be both resource and time consuming (as forewarned by Jordaan, 1989) and the development of original tests remains, for the moment, a developing area.

In the absence of the full development of South African English speech discrimination test, South African audiologists have traditionally used the United States of America's Central Institute for the Deaf (C.I.D.) W-22 wordlists, presented using monitored live voice. The use of non-South African speech and speech related tests, both in their original form and in modified forms, has occurred

in South Africa for many years with varying degrees of success. Examples include (amongst others): the WIPI and NU-CHIPS speech discrimination tests (Mehl, 1992), the Staggered Spondaic Word test (Collie, 1991; Stevens, 1991; Corke, 1993; Modi, 1995), the Paediatric Speech Intellegibility test (Ewen, 1993), the Willeford Battery of Central Auditory Function (Kastner, 1994; Modi, 1995), and the Peabody Picture Vocabulary Test-Revised (Alant & Beukes, 1986; Labuschagne 1990; Koekemoer, 1994; Pakendorf & Alant, 1997). Whilst the reliability and validity of using these tests on South African speakers has been formally assessed, the use of the C.I.D. W-22 on South African English speakers appears to be due mainly to its high and well-documented acceptance in the U.S.A only, and not because of any published data validating its use on the South African population.

The use of the C.I.D. W-22 in South Africa introduces many reliability and validity problems. Firstly, there is the effect of familiarity and context (or expectation). Familiar words (Ostergard, 1983) and phonemes (Lehiste & Peterson, 1959) can be identified up to 15 dB lower than

those that are unfamiliar (Edgerton & Danhauer, 1979). Secondly, an ideal speech tests should contain an equal phonetic composition representative of everyday speech in the test population (Egan, 1948). In applying the C.I.D. W-22 to the South African population, any equal phonetic representation is lost. Tobias (1964) and Campbell (1965) questioned the need for equal phonetic representation, however, by claiming there was overwhelming clinical and experimental evidence that this an interesting, but unnecessary component in speech audiometry testing. Even today, experimental evidence is lacking and the issue remains unresolved (Mendel & Danhauer, 1997). Finally, a valid speech discrimination test should contain an equal range and average level of difficulty within and between wordlists. Although this ideal is rarely achieved in reality (Ross & Huntington, 1962), large variations will affect test reliability and sensitivity. These factors all strongly influence, and possibly invalidate, the use of the C.I.D. W-22 with its American English normative data on the South African English speaking population.

In response to the problems faced by the C.I.D. W-22, many South African audiologists make use of monitored live voice (MLV) presentation. Whilst MLV provides a partial solution to speaker/listener accent mismatch, it does not reduce the dialectal differences inherent in the test materials themselves. Added to this is MLV's well-recognised poor test-retest reliability (Carhart, 1965; Brandy, 1966; Kreul, Bell & Nixon, 1969; Northern & Hattler, 1974), and the problem of comparing MLV test results to standardisation information obtained using recorded versions of the stimuli (Mendel & Danhauer, 1997). Considering the problems already faced in using non-South African English speech material, the significantly greater variability of MLV could effectively negate any advantages it may offer in overcoming accent differences.

If MLV is to be avoided (an opinion not held by all clinicians), this returns the South African audiologist to the option of using the C.I.D. W-22 in its pre-recorded form on cassette tape or compact disc. In addition to the previously mentioned problems this introduces, closer inspection of the C.I.D. W-22 reveals that it may not be the best option for the South African English speaker in the first place.

The C.I.D. W-22 was designed as an improvement on one of the first pre-recorded speech tests, the Psycho-acoustic Laboratories phonetically-balanced 50-word lists (PAL PB-50). It improved phonetic balance (Lehiste & Peterson, 1959), equality of difficulty between lists (Brewer & Resnick, 1983) and degree of familiarity (Brewer & Resnick, 1983; Hirsch et al., 1952). The test consists of 4 lists of 50 monosyllabic words with no repetitions. The words were chosen to be familiar to North American English speakers, and the phonetic composition of each list aims to reflect that of English as a whole (still based, however, on North American English). Each word is preceded by a carrier phrase and each list may be presented with one of six different word orders. In its recorded form, it is spoken by an adult male with "general North American" accented English and all its normative values are based on North American English speakers (Hirsch et al., 1952).

Despite the improvements shown by the C.I.D. W-22, it still has several faults that will only be worsened when applying the test to the South African English speaking

population. Brewer and Resnick (1983) and Ostergard (1983) question the C.I.D. W-22's ability to discriminate effectively among the performance of the hearing impaired, claiming that it is too easy. Ross and Huntington (1962) and Studebaker and Sherbecoe (1991) disclaim equality of difficulty between lists. Mendel and Danhauer (1997) have also claimed that the validity and reliability of the test for certain conditions under which it is used, such as in background noise, are poorly documented. Finally, users recognised almost immediately the benefits of reducing the test-time (Edgerton & Danhauer, 1979) by administering 25 word half-lists (Martin & Forbis, 1978). However, these shorter lists are prone to higher variability and therefore poorer reliability and are no longer phonetically balanced (Ostergard, 1983). Even considering Cakiroglu and Danhauer's (1992) claim that listener's and talker's linguistic background has statistical, but not clinical, effects on the performance of non-native English-speaking subjects on the C.I.D. W-22 (although they do warn that listener's having less experience with English may not perform as well as native-English speakers), these identified problems warrant a search for a better speech discrimination test for the South African English speaker.

An alternative which claims to combat many of the problems encountered by whole word tests, such as the C.I.D. W-22, is the phonemically scored speech test (Edgerton & Danhauer, 1979). Lehiste and Peterson (1959) argued that as the speech signal is not acoustically invariant, it is impossible to achieve phonetic balance in a speech test. Therefore intelligibility (discrimination) measures should use materials that are phonemically balanced [although again, the need for phonemic balancing in the first place has been queried (Mendel & Danhauer, 1997)].

Phonemic scoring offers advantages of general and specific significance to a South African English speaking population. Generally, factors such as increased reliability, abbreviated test lists, and decreased test time (Edgerton & Danhauer, 1979) are all beneficial. Specifically, factors such as a decrease in the effect of the listener's familiarity with the vocabulary used, and the controversial claim that phoneme scoring involves less extrinsic redundancy and therefore eliminates the effects of linguistic bias (Lyregaard, 1987; contested by Keith, Katbamna, Tawfik & Smolak, 1987), suggest that use of normative values supplied with a phonemically scored wordlist could provide greater test reliability and validity when used on South African English speakers.

Following the success of the first phonemically scored wordlists developed in the U.S.A. by Lehiste and Peterson in 1959 (the consonant-nucleus-consonant or CNC lists), and Tillman, Cahart and Wilber in 1963 (the Northwestern University Auditory Tests No.4 and No. 6), Arthur Boothroyd of the United Kingdom developed the phonemically scored Arthur Boothroyd (AB) wordlists (Boothroyd, 1968). Boothroyd's test contains 15 isophonemic lists, each containing 10 words with 30 scorable phonemes. Each list contains the same 20 consonants and 10 vowels, in varying combinations (Dermody & Mackie, 1987). The phonemes chosen were those found to be most frequent in Boothroyd's vocabulary, and the only limitations placed on the original word lists were that neither first names nor obscenities would be included (Boothroyd, 1968).

The AB wordlists received criticism, however, with suggestions that Boothroyd's Northern English accent

impacted on the design of the test's phoneme content, and that the lists did not comprise random samples of phonemes. Despite Boothroyd's reply that the phonemes are altered by changing their word position, and therefore the phoneme arrangements are essentially random (which would appear to undermine his claim of equality of phonetic representation), the AB wordlists were recorded on at least 2 occasions in the United Kingdom using different U.K. accents (Evans, 1987).

The success of the phonemically scored AB lists in the U.K. led to their use in Australia by the National Acoustic Laboratories (NAL) (Dermody & Mackie, 1987). NAL produced a revised version specifically for Australian use, with rearranged word orders and several new words to replace those considered to have low familiarity for Australian English speakers (Travers, 1990). According to Bench (1987), Australian English is based on that of the United Kingdom, but includes North American influences and some uniquely Australian expressions, while some words common to both dialects may carry different meanings. The NAL-AB wordlists (appendix one) quickly became the most commonly used speech discrimination test in Australia (Bench, 1987), resulting mainly from advantages such as short administration time, and the availability of a high number of lists allowing multiple testing to occur in the one session without learning effects or fatigue (Markides, 1987; Mendel & Danhauer, 1997).

In view of the success of the pre-recorded AB wordlists in the U.K. and Australia, the possibility that they could offer significant advantages over the present MLV presentation of the C.I.D. W-22 wordlists in South Africa becomes apparent. The problem of linguistic bias remains, however, and despite claims that phonemic scoring decreases the effect of linguistic bias (Lyregaard, 1987; contested by Keith, Katbamna, Tawfik & Smolak, 1987), and can be used even on clients with limited familiarity with English (Evans, 1987), it is vital to consider the validity of using this test in a multilingual population such as South Africa's.

AIMS

The aims of this study were as follows:

- 1) To determine if first language South African English speakers performed differently to first language, Australian English speakers, on the pre-recorded NAL-AB wordlists.
- 2) To determine if the first ten (of fifteen) pre-recorded NAL-AB wordlists were of equal difficulty for first-language, South African English speakers.
- 3) To investigate the feasibility of using the pre-recorded NAL-AB wordlists, with the NAL Australian English normative data, to assess the speech discrimination abilities of first-language South African English speakers in South Africa.

METHODOLOGY

Exact replication of the NAL-AB wordlist normative study proved to be difficult. The normative data on the performance of Australian English speakers on the pre-recorded NAL-AB wordlists is documented in a clinical manual entitled "AB Word Lists: NAL Protocols" prepared by Anne Travers, NAL, Victoria, August, 1990. This

document is supplied with the NAL-AB Word List compact disk and outlines graphically all normative data for the 15 wordlists separately, and as a total average, and contains full recommended procedures the test's clinical use. On close inspection, however, the document provides little detail as to the exact procedures used to obtain the normative values. Details of the NAL-AB wordlist study were not published in the literature and the original data has since been removed from NAL archives. The only fact that is clearly stated in the Travers document is that the data was obtained using 120 normally hearing subjects. With this limited information in mind, the following methodology was used.

SUBJECTS

Subject selection criteria

Subjects were selected using a convenience sampling technique according to the following criteria: (i) Subjects should be adults between 17 and 40 years. This controlled for the well-recognised effects of paediatric and geriatric age ranges on performance on speech discrimination tasks (Hall, 1983). (ii) Subjects should be female. This criterion was included for ease of subject selection, as it was deemed unlikely that a sample balanced for gender could be easily obtained. (iii) Subjects should be first-language speakers of South African English. (iv) Subjects should be 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) As determined by a pretest audiogram, the test ear should have hearing thresholds no greater than 5dBHL at 250, 500, 1000, 2000, 4000 and 8000 Hz. (v) As determined by self-report on a pretest questionnaire, subjects should have no significant history, past or present, of: hearing impairment; speech or language impairment; tinnitus; ear infections; noise exposure; speech/articulation disorders; the presence of any disorders of language or; family history of hearing problems. (vi) Subjects should have no knowledge or experience with the NAL-AB wordlists specifically. (vii) Subjects should have little or no knowledge or experience of Australian English.

Subject description

An initial sample of 32 female, first language South African English speaking subjects, all resident in the Gauteng region was collected. Two subjects were excluded when pure tone screening revealed unacceptable hearing thresholds leaving a final 30 subjects ranging in age from 17.5 years to 26.0 years, with a mean age of 20.5 years, a standard deviation of 1.8 years, and a median age of 21.0 years. Education levels included two subjects with secondary level, and 28 with, or receiving, tertiary level education. 26 of the 28 tertiary level subjects were studying in the field of speech pathology and audiology. On the basis of selecting the best hearing ear as the test ear, 12 right and 18 left ears were tested.

TEST ENVIRONMENT, EQUIPMENT AND RECORDED MATERIAL

Testing took place in a two-room sound treated booth at the University of the Witwatersrand Hearing Clinic in

Johannesburg, Gauteng, South Africa. The booth included a one-way mirror to allow the tester to observe the subjects during testing. The NAL-AB wordlists were presented via a Sanyo compact disc player CDP-41 coupled with a Grayson-Stadler GSI16 audiometer. Presentation was by means of TDH-50 headphones with MXAR41 cushions. The NAL Speech Recognition Materials (Disc 1) compact disc with intensity levels equated using an A-weighted Long Term Equivalent Level (Leq) procedure provided the speech stimuli.

The NAL-AB Wordlist CD provides 15 pre-recorded AB wordlists (appendix one) arranged to be as representative as possible of 'typical' Australian English (Travers, 1990). Lists are spoken by an adult male Australian general English speaker. Each list contains the same 20 consonants and 10 vowels, arranged into various combinations of 10 CVC monosyllables per list. The lists are ordered according to their reliability as measured on the Australian population (Dermody, cited in Travers, 1990). The order reflects a pairing of those lists with the most similar performance-intensity functions. The last three of the 15 lists demonstrate the most widely variant functions and NAL recommends that these be used at the top end of the PI function only. Each of the word lists is preceded by a short warning pip, the inter-stimulus interval is 4 seconds which serves as the subject response period (there is no carrier phrase), and each list is recorded on a separate track.

MEASUREMENT PROCEDURES AND DATA COLLECTION

On passing the initial hearing and case history questionnaire screening, subjects were assessed on their speech discrimination performance on the first 10 of the 15 pre-recorded NAL-AB wordlists in their better hearing ear. Administration of all 15 lists to both ears was considered to be impracticable for 3 reasons: (1) the duration of testing would be increased to unacceptably high levels possibly leading to subject fatigue; (2) the additional repetitions of each list would induce a greater degree of practise/learning effects; and (3) as lists 13-15 were shown to be more variable in the Australian sample (Dermody, cited in Travers, 1990), it was felt that lists 1-10, the most reliable and commonly used lists, were more pertinent to the aims of this study.

An identical set of instructions was given to all subjects, via the headphones, as per NAL recommendations (Travers, 1990, p8). The instructions read as: "You are going to hear some single words. After each word there will be a pause for you to repeat the word. Listen carefully and repeat whatever you hear. Some of the words will be very soft and it is important that you have a guess at each word even if it doesn't make sense or you only hear a part of the word such as 'ee' or 'ch'. Just repeat what you hear." Once subjects had been instructed, list 11 was presented at 30dBHL (audiometer dial reading), in order to familiarise subjects with the test procedure.

The wordlist presentation was pseudorandomised with the initial ten list order being randomly selected, and then repeated five times. This was done to prevent a particular list from occurring several times in succession, or in close proximity. The intensity presentation was then randomised throughout all 50 presentations such that each of the ten lists was presented once at five different intensity levels.

The whole process of pseudorandomisation was conducted separately for each individual subject. Intensity levels of 0, 5, 10, 20 and 25 dBHL (as per the audiometer dial reading) were used, emulating the normative data provided by NAL.

All subject's responses were recorded and scored on-line by a single tester, a student of Speech and Hearing Therapy at the University of the Witwatersrand, under the supervision of a full member of the Audiological Society of Australia, in fulfilment of that Society's requirements for use of materials. Responses were transcribed using a combination of orthographic and phonetic transcription. Each word was then assigned a score of 0%, if no phonemes were correctly repeated; 3% for 1 correct phoneme; 7% for 2 correct phonemes; or 10% if the response was completely correct; as per the scoring protocols suggested by NAL (Travers, 1990).

Subject responses were considered to be correct only if they were pronounced exactly as per the Australian English pronunciation on the NAL-AB wordlist CD. It was noted, however, that none of the 15 NAL-AB wordlists contained words where a predicted difference in pronunciation between Australian English and South African English should occur (for e.g., there were no words such as "dance", which would be pronounced as /dæns/ in Australian English, but as /dans/ in South African English). In cases where single phonemes or nonsense syllables were given as a response instead of a whole word, the phonemes were scored based on their acoustic similarity to the relevant Australian English phonemes in the stimulus word. In the case of more than one response being given for a single stimulus, the first response was taken irrespective of accuracy. Subjects were allowed the opportunity to rest at any stage during the test procedure.

DATA ANALYSIS

Comparison between the South African English speaker scores and the NAL-AB wordlist Australian English normative data was done graphically with all NAL normative data values being extrapolated from the graphs provided in Travers (1990) for reanalysis. Plots of the extrapolated NAL normative data and the South African English speakers data were completed using Microsoft Excel for Windows, Version 5.0®.

Analysis for differences within the South African English speaker discrimination scores, between presentation intensities, was conducted separately for each wordlist using Kruskal-Wallis ANOVA by Ranks (across all intensities at $p < 0.05$) and Mann-Whitney U (between adjacent intensities only, at $p < 0.01$ to allow for the multiple comparisons) analyses. Non-parametric statistics were used as a result of the non-homogeneity of variances (determined by the F test at the $p < 0.01$ level) and the significant correlation of mean scores to variances present in the data (determined by the Pearson's Product Moment Correlation Coefficient test at the $p < 0.05$ level).

Analysis for differences within the South African English speaker discrimination scores, between wordlists, was conducted separately for each presentation intensity using one way, repeated measures ANOVA and Tukey Honest Significant Difference analyses (chosen as it is one of the more conservative of the post hoc analysis tests) at the $p < 0.05$ level.

Analyses for correlations in the South African English

speaker discrimination scores between wordlists was conducted separately for each presentation intensity using Pearson's Product Moment Correlation Co-efficient analyses (at the $p < 0.05$ level). Similar analyses for correlations between presentation intensities was also conducted separately for each wordlist.

All statistical analysis was conducted using Statistica for Windows, Version 4.5 software from Statsoft Inc, 1993[®].

RESULTS

SOUTH AFRICAN ENGLISH SPEAKERS' VERSUS AUSTRALIAN ENGLISH SPEAKERS' RESULTS

Figure 1 shows the South African English speakers' speech discrimination results, superimposed on the extrapolated NAL Australian English normative data, for each individual NAL-AB word list. Overall, the NAL Australian English speakers' mean scores tended to be better than the South African English speakers' mean scores with greater differences occurring at the lower presentation intensities. Extrapolated mean values for the NAL Australian English normative data generally fell within 1 standard deviation above the South African English speakers' means. Only list four at 5 dBHL, list eight at 0 dBHL and list nine at 5 dBHL, showed the NAL Australian English scores to be greater than one standard deviation above the mean South African English scores on visual inspection.

Specific comparison between the mean South African English speakers' speech discrimination scores and the extrapolated NAL Australian English normative data showed three basic (qualitative) categories:

- 1) Similar mean scores across four or more presentation levels (lists one and five).
- 2) South African English mean scores falling below the NAL Australian English mean scores in the mid and/or low presentation levels only (lists three, four, seven, eight and nine).
- 3) South African English mean scores falling below the NAL Australian English mean scores across four or more presentation levels (lists two, six and ten).

COMPARISONS WITHIN SOUTH AFRICAN ENGLISH SPEAKERS' RESULTS

Table 1 shows the mean and standard deviation speech discrimination scores for the 30 South African English

speakers on all ten NAL-AB wordlists, at all five presentation intensities. Table 2 shows Kruskal-Wallis ANOVA by Ranks and Mann-Whitney U analysis results for differences in the scores between presentation intensities, conducted separately for each wordlist. Significant differences ($p < 0.01$) were shown between all adjacent presentation intensities in each wordlist except between 20 and 25 dBHL for lists 1, 2, 5, and 6. Tables 3 and 4 show one-way repeated measures ANOVA and Tukey Honest Significant Difference test results for differences in scores between wordlists, conducted separately for each presentation intensity. Results showed multiple significant differences between wordlists at all five presentation intensities, but predominantly at the higher intensities.

Pearson's Product Moment correlation co-efficient analyses ($p < 0.05$) between scores at each presentation intensity, conducted separately for each wordlist, showed mild (0.4) to strong (0.9) correlations between intensities for all wordlists. A small number (<10% of the total number of correlations) of poor (<0.4) or non-significant ($p > 0.05$) correlations occurred within the wordlists between low and high presentation intensities only. Similar analyses between scores across wordlists, conducted separately for each presentation intensity, showed mild (0.4) to strong (0.9) correlations between scores for all presentation intensities except for 25 dBHL, which showed large numbers (24.4%) of poor (<0.4) or non-significant ($p > 0.05$) correlations between wordlists.

F Test analysis for differences in the variance of the South African English speakers' discrimination scores between the different presentation intensities, conducted separately for each wordlist, showed the variance at 25 dBHL to be significantly different ($p < 0.01$) to the variance at 0, 5 and 10 dBHL in lists 1, 2, 3, 4, 6, 9 and 10; to the variance at 0 and 5 dBHL in list 5; and to the variance at 5 and 10 dBHL in list 8. The variance at 20 dBHL was shown to be significantly different ($p < 0.01$) to the variance at 0 and 5 dBHL in list 3 and to the variance at 0 dBHL in list 5. List 7 showed no significant differences ($p < 0.01$) in variance at any presentation intensity.

DISCUSSION

The aims of this study were: (i) to determine if first language, South African English speakers performed differently to Australian English speakers on the pre-recorded NAL-AB wordlists, (ii) to determine if the pre-recorded NAL-AB wordlists were of equal difficulty for first-language, South African English speakers and (iii)

TABLE 1: South African English speakers' mean and S.D speech discrimination scores on NAL-AB wordlists 1-10.

dBHL	0	5	10	20	25
LIST 1	24.8 ± 18.1	52.5 ± 22.4	69.1 ± 18.4	91.0 ± 10.2	95.6 ± 5.7
LIST 2	23.2 ± 18.1	43.8 ± 19.3	63.0 ± 18.7	85.7 ± 10.6	90.1 ± 5.9
LIST 3	27.2 ± 24.6	47.1 ± 22.1	65.8 ± 18.3	87.9 ± 8.3	92.2 ± 6.8
LIST 4	19.9 ± 15.1	37.6 ± 16.5	60.2 ± 19.1	89.3 ± 10.5	96.3 ± 5.7
LIST 5	29.5 ± 22.9	50.2 ± 20.2	69.7 ± 15.0	90.0 ± 9.0	93.5 ± 6.5
LIST 6	23.3 ± 18.2	47.6 ± 20.5	66.9 ± 18.1	84.4 ± 12.4	90.6 ± 7.5
LIST 7	20.6 ± 17.2	40.7 ± 18.6	58.7 ± 18.0	86.9 ± 11.0	92.5 ± 9.8
LIST 8	18.2 ± 14.2	42.1 ± 17.4	62.1 ± 17.1	83.1 ± 8.4	89.1 ± 6.5
LIST 9	23.8 ± 18.9	47.0 ± 19.0	66.2 ± 18.2	88.8 ± 11.0	94.3 ± 7.3
LIST 10	19.8 ± 16.1	45.4 ± 18.7	57.0 ± 19.6	82.8 ± 10.2	88.4 ± 6.4

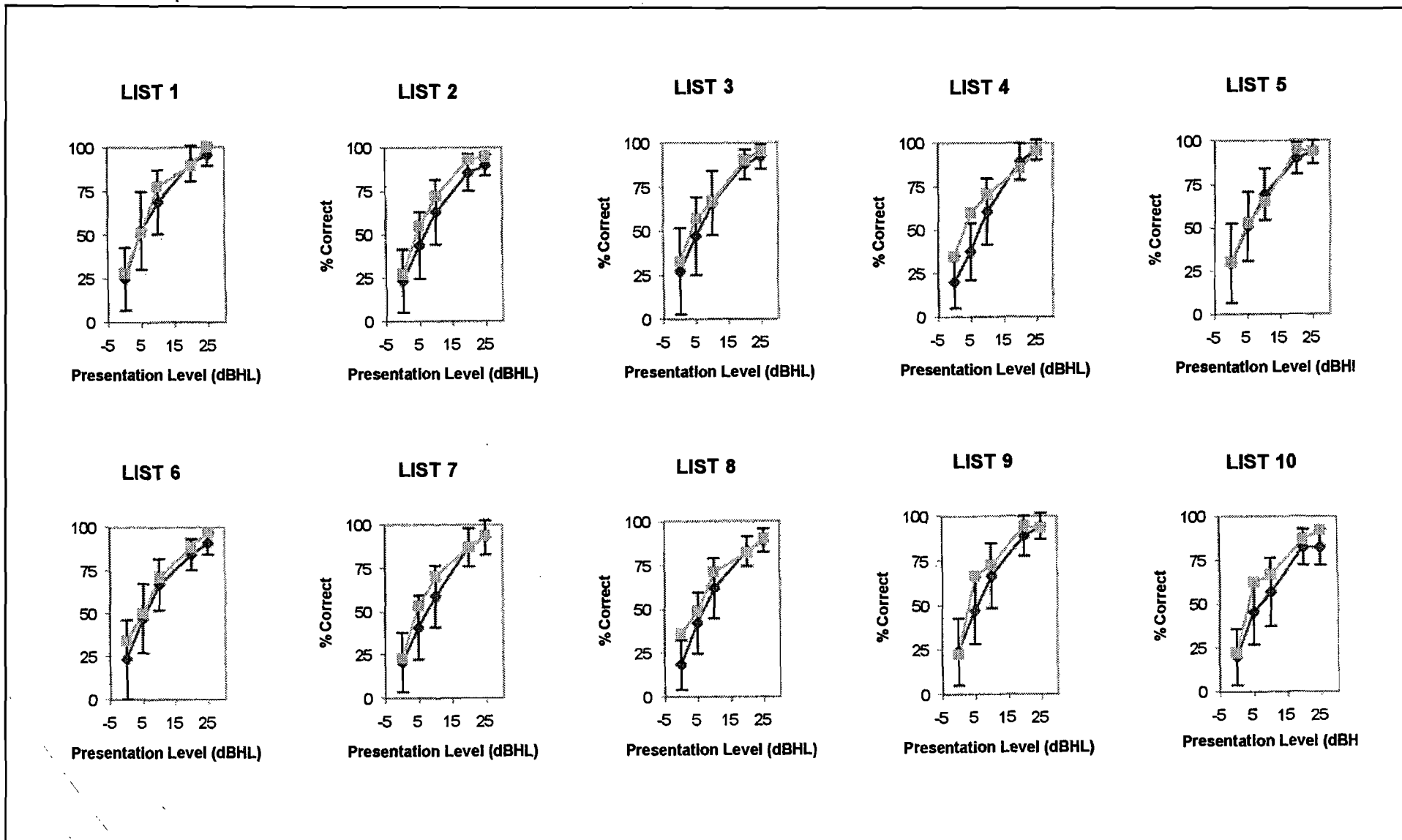


FIGURE 1: Plots of South African English speakers' mean scores (diamonds) with error bars (1 S.D.) and NAL's Australian English speaker mean scores (squares) for NAL-AB wordlists 1-10.

to investigate the feasibility of using the pre-recorded NAL-AB wordlists, with the NAL-Australian English normative data, to assess the speech discrimination abilities of first-language South African English speakers in South Africa.

To achieve these aims, NAL-AB wordlist speech discrimination scores were obtained from 30 first language South African English speaking, female adults with normal hearing. These scores were then compared to the NAL Australian English speaker normative data (as provided by Dermody, in Travers, 1991). In general, the South African scores were lower, with three qualitative patterns being observed; similar performance between the two groups (lists one and five); South African scores occurring below the Australian scores at the lower and/or mid presentation levels only (lists three, four, seven, eight and nine); and South African scores occurring below the Australian scores across four or all five presentation intensities (lists two, six and ten). Comparisons of South African scores within and between NAL-AB wordlists showed multiple differences and

correlations, depending on which lists and presentation intensities were being compared.

The qualitative finding that South African scores were generally worse than those of the Australians, indicates the direct use of the pre-recorded NAL-AB wordlists with their associated normative values cannot be applied to the South African English speaking population without significant modification.

The three differing patterns of South African versus Australian results indicates the different NAL-AB wordlists were not of equal difficulty for the two populations. The first category of results showing similar mean scores for the two groups, suggests these lists were equally difficult. The second category showing differences at the low and/or mid-presentation intensities only, suggests both groups were able to hear these wordlists equally well at the higher presentation intensities, but the Australians were able to extract more information at the lower presentation intensities, due most probably to their

TABLE 2: Kruskal-Wallis ANOVA by Ranks and Mann-Whitney U results for differences in South African English speakers' mean scores between presentation intensities, conducted for each NAL-AB wordlist separately (*-highly significant).

NAL-AB WORDLIST NUMBER	KRUSKAL-WALLIS ANOVA BETWEEN THE FIVE PRESENTATION INTENSITIES FOR EACH WORDLIST		Mann-Whitney U (n=60) results between adjacent presentation levels for worldlist
	n	p level	
1,2,3,4,5,6,7,8,9,10	150	<0.005*	Sig. diff's (p<0.01*) between all adjacent presentation levels except between 20 and 25 dBHL on lists 1, 2, 5, and 6.

TABLE 3: Repeated measures, one way ANOVA results for differences in South African English speakers' mean scores between NAL-AB wordlists, conducted for each presentation intensity separately (*-highly significant).

PRESENTATION INTENSITY (dBHL)	Degrees of Freedom (10 NAL-AB wordlists)	p level
0,5,10,20,25	9	<0.001*

TABLE 4: Results of Tukey analysis for differences (p<0.05) in South African English speakers' mean scores between wordlists when presented at the same presentation intensity, e.g., lists 8 and 5 showed significantly different scores when presented at 0, at 20, and at 25 dBHL, but not at 5 or at 10 dBHL.

NAL-AB List No.	1	2	3	4	5	6	7	8	9	10
1		20,25		5		20,25	5,10	20,25		10,20,25
2				25					25	
3								20		10,20
4					0,5,10	25		10,20,25		10,20,25
5						20	0,10	0,20,25		0,10,20,25
6										10
7										
8									20,25	
9										10,20,25
10										

familiarity with Australian English. The third category showing a poorer South African English speaker performance across four or all five presentation intensities, suggests that even the advantage of increased presentation intensity did not allow the South Africans to overcome the difficulties encountered in listening to Australian English.

This "three category ranking" of Australian versus South African scores places the ten NAL-AB wordlists in a "valid" (category one) through to "use with caution" (category three) ranking for South African English speakers. This supports Markide's 1987 disagreement with Boothroyd's 1968 and Keith's 1987 claims that phoneme scored speech discrimination tests generally, and the AB wordlists specifically, can be used accurately with non-English speakers. It also questions Evans' 1987 suggestion that subjects with minimal knowledge of English may be tested with AB lists by a clinician whose own experience with the test is extensive. Evans stated the PI functions are likely to be less steep, approximating those for nonsense syllables, but maximum scores may still reach 100%. This study showed equal or steeper sloped performance intensity functions for the South African English speakers, with the steeper functions seen in category two.

The 4:1-Australian:South African ratio of sample sizes used in this study prevents any comment on whether there was a true difference in the variance of the speech discrimination scores between the Australians and South Africans. A notable observation, however, was the significant decrease in the variance of the South African scores as the presentation intensity increased. This finding appears to be similar (on visual inspection only) to the Australian plots (Travers, 1990), and only partly agrees with Ostergard's (1983) findings that variability in speech discrimination scores tend to decrease for extreme scores and increase for midrange scores.

Difference analyses within the South African scores themselves, further indicated that the NAL-AB wordlists were not of equal range and difficulty for the South African English speakers (although a similar conclusion was found for the Australian English speakers by NAL themselves). Overall, the multiple significant differences seen indicate that the performance of the South Africans was dependent on which wordlist and presentation intensity was being used. Interpretation of the significant differences must be approached with caution, however, as close inspection shows many of these to be more dependent on differences in variance, rather than any true clinical differences in mean scores. Even so, the South Africans did show a steady improvement in scores on all wordlists with increases in presentation intensity, with evidence of the beginning of a ceiling effect at the 25 dBHL level.

Correlational analyses within the South African scores showed a high number of significant positive correlations between different lists at the different presentation intensities, particularly for the mid-presentation intensities. This indicates a good relationship between NAL-AB wordlists for the South Africans in that a high score on one wordlist at one intensity, will generally predict a high score on the same or another wordlist at the same or different intensity (and vice versa). The exception to this was at 25 dBHL where a large number (24.4%) of poor (<0.4) or non-significant ($p>0.05$) correlations occurred between wordlists. This finding supports the earlier suggestion of the beginning of a ceiling at 25 dBHL presentation intensity as scores in a ceiling range should

be more random and therefore have fewer correlations.

Minor limitations to this study include the all 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 the sample's English first language prerequisite, and predominance of tertiary level education, much of which was in the field of Speech Pathology and Audiology. 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 a certain level of linguistic competence and even sophistication. It might therefore be predicted that a more representative sample of the South African English 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 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, Thorton 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 be. 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 Australian English and South African English speakers on the pre-recorded NAL-AB wordlists makes this test a valid option for the speech discrimination assessment of South African English speakers, under certain conditions. The NAL-AB wordlists are most suitable for use at suprathreshold levels (where the fewest differences were observed) where the advantages of shorter test times, and more available wordlists, can be accessed without critically compromising test reliability or validity. Use of the NAL-AB wordlists at threshold, or near threshold, levels (where the most differences were observed) should be approached with caution, however, and reliance on these wordlists for site of lesion purposes should be avoided. General test reliability and validity could be improved by changing the wordlist presentation order to (5, 1), (3, 9), (7, 6), (8, 4), (2, 10) (where brackets indicate interchangeable lists) to suit South African English speakers, and by collecting larger South African normative data bases where possible.

Despite this study's findings, the over-riding need for a South African English specific speech discrimination test remains. Whilst the NAL-AB wordlists offer advantages over the C.I.D. W-22 wordlists such as phonemic scoring and, based on this study's results, at least some idea of how South African English speakers vary from the normative data (information which is not currently published for the C.I.D. W-22 wordlists), the continued use of non-South African tests must be seen as an interim measure only.

In view of similarities between the performance of South

African English and Australian English speakers seen in this study, the option of modifying the NAL-AB wordlists to suit South African English (as per NAL's modification of the original U.K. AB wordlists to suit Australian English) could provide the starting point needed to develop the long awaited South African English specific speech discrimination test.

ACKNOWLEDGEMENT

The authors wish to thank Nicci Campbell of the Department of Communication Pathology, University of Pretoria, for her work in finding many of the South African studies referenced in this paper.

REFERENCES

- Alant, E. & Beukes, S.M. (1986). The application of the revised version of the Peabody Picture Vocabulary Test (PPVT-R) to non-mainstream children. *South African Journal of Communication Disorders*, 33: 7-12.
- Bench, J. (1987). Speech Audiometry in Australia. In M. Martin (Ed.), *Speech Audiometry*. London: Whurr Publishers Ltd.
- Boothroyd, A. (1968). Developments in speech audiometry. *Sound*, 2: 2-10.
- 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.
- Cakiroglu, S. & Danhauer, J.L. (1992). Effects of listeners' and talkers' linguistic backgrounds on W-22 test performance. *Journal of the American Academy of Audiology*, 3: 186-192.
- Campbell, R.A. (1965). Discrimination test word difficulty. *Journal of Speech and Hearing Research*, 8: 13-22.
- Carhart, R. (1965). Problems in the measurement of speech discrimination. *Archives of Otolaryngology*, 82(9): 253-260.
- Collie, L. (1991). *The Staggered Spondaic Word (SSW) Test investigated in a South African Context*. Unpublished undergraduate research report, Department of Speech Pathology and Audiology, University of the Witwatersrand.
- Corke, V. (1993). *The Staggered Spondaic Word Test: The South African English version*. Unpublished undergraduate research report, Department of Speech Pathology and Audiology, University of the Witwatersrand.
- Davis, A.C. (1983). Hearing disorders in the population.. First phase findings of the MRC National Study of Hearing. In M.E. Lutman & M.P. Haggard (Eds.), *Hearing Science and Hearing Disorders*, London: Academic Press.
- Dermody, P. & Mackie, J. (1987). Speech tests in audiological assessment at the National Acoustic Laboratories. In M. Martin (Ed.), *Speech Audiometry*. London: Taylor and Francis Ltd.
- Edgerton, B.J. & Danhauer, J.L. (1979). *Clinical Implications of Speech Discrimination Testing Using Nonsense Stimuli*. Baltimore: University Park Press.
- Egan, J.P. (1948). Articulation testing methods. *Laryngoscope*, 58: 955-991.
- Evans, P.I.P. (1987). Speech audiometry for differential diagnosis. In M. Martin (Ed.), *Speech Audiometry*. London: Taylor and Francis Ltd.
- Ewen, C. (1993). *The Paediatric Speech Intelligibility Test-Format II: Performance-intensity functions of first language English speaking South African children*. Unpublished undergraduate research report, Department of Speech Pathology and Audiology, University of the Witwatersrand.
- 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. III (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.
- Jordaan, H. (1989). Management of language disability in the bilingual, bicultural child. In *SAALD Conference Proceedings: Facilitating Learning: An Ecological Perspective*, July: 217-234.
- Kastner, C. (1994). *The Willeford Battery of Central Auditory Function: The South African English version*. Unpublished undergraduate research report, Department of Speech Pathology and Audiology, University of the Witwatersrand.
- Keith, R.W., Katbamna, B., Tawfik, S. & Smolak, L.H. (1987). The effect of linguistic background on staggered spondaic word and dichotic consonant vowel scores. *British Journal of Audiology*, 21: 21-26.
- Koekemoer, H. (1994). *Die toepaslikheid van 'n Noord-Sotho vertaling van die PPVT-R by 'n groep skoolgaande kinders*. Unpublished undergraduate research report, Department of Communication Pathology, University of Pretoria.
- 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.
- Labuschagne, D.A. (1990). *The application of the PPVT-R to an urban group of North-Sotho speaking children*. Unpublished undergraduate research report, Department of Communication Pathology, University of Pretoria.
- Lehiste, I. & Peterson, G.E. (1959). Linguistic considerations in the study of speech intelligibility. *Journal of the Acoustic Society of America*, 31: 280-286.
- Lyregaard, P. (1987). Towards a theory of speech audiometry tests. In M. Martin (Ed.), *Speech Audiometry*. London: Taylor and Francis Ltd.
- Markides, A. (1987). Speech tests of hearing for children. In M. Martin (Ed.), *Speech Audiometry*. London: Taylor and Francis Ltd.
- Martin, F.N. & Forbis, N.K. (1978). The present status of audiometric practice: A follow-up study. *ASHA*, 20: 531-541.
- Mehl, T. (1992). *Speech audiometry testing for pre-school first language English speaking South African children: Comparison of the WIPI and NU-CHIPS speech discrimination tests*. Unpublished undergraduate research report, Department of Speech Pathology and Audiology, University of the Witwatersrand.
- Mendel, L.L. & Danhauer, J.L. (1997). *Audiological Evaluation and Management and Speech Perception Assessment*. San Diego: Singular Publishing Group Inc.
- Modi, N. (1995). *Use of the SA-SSW and the SA-Willeford battery with South African Indian speakers of English*. Unpublished undergraduate research report, Department of Speech Pathology and Audiology, University of the Witwatersrand.
- Northern, J.L. & Hattler, K.W. (1974). Evaluation of four speech discrimination test procedures on hearing impaired patients. *Journal of Auditory Research*, (Suppl): 1-37.
- Ostergard, C.A. (1983). Factors influencing the validity and reliability of speech audiometry. *Seminars in Hearing*, 4: 221-240.
- Pakendorf, C. & Alant, E. (1997). Culturally valid assessment tools: Northern Sotho translation of the Peabody Picture Vocabulary Test-Revised. *South African Journal of Communication Disorders*, 44: 3-13.
- Ross, M. & Huntington, D.A. (1962). Concerning the reliability and equivalency of the C.I.D. W-22 auditory tests. *Journal of Auditory Research*, 2: 220-228.
- Stevens, P. (1991). *The effect of linguistic background on the Staggered Spondaic Word (SSW) test performance*. Unpublished undergraduate research report, Department of Speech Pathology and Audiology, University of the Witwatersrand.
- Studebaker, G.A. & Sherbecoe, R.L. (1991). Frequency-importance and transfer functions for recorded C.I.D. W-22 word lists. *Journal of Speech and Hearing Research*, 34: 427-438.
- Taylor, O.L. (1986). *Treatment of communication disorders in linguistically diverse populations*. Boston: College-Hill.
- 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.
- Tillman, T.W., Cahart, R. & Wilber, L. (1963). *A test for speech discrimination composed of CNC monosyllabic words*. Northwestern University Auditory Test No. 4. Technical Documentary Report No. SAM-TDR-62-135, USAF School of Aerospace Medicine, Brooks Airforce Base, Texas.
- Tobias, J.V. (1964). On phonemic analysis of speech discrimination. *Journal of Speech and Hearing Research*, 7: 98-100.
- Travers, A. (1990). AB Word Lists: NAL Protocols. Australian Hearing Services, Victoria, Australia.

APPENDIX 1

National Acoustics Laboratories Arthur Boothroyd Wordlists one to fifteen.

List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	List 9	List 10
fan	fun	thug	hush	jug	bath	have	hug	math	wish
rug	will	witch	gas	latch	hum	wig	dish	hip	dutch
ship	vat	teak	thin	wick	dig	buff	ban	gun	jam
cheek	shape	rap	fake	faith	five	mice	rage	ride	heath
haze	wreath	vice	chime	sign	ways	teeth	chief	siege	laze
dice	hide	jail	weave	beep	reach	jays	pies	veil	bike
both	guess	hen	jet	hem	joke	poach	wet	chose	rove
well	comb	shows	rob	rod	noose	rule	cove	shoot	pet
jot	choose	food	dope	vote	pot	den	loose	web	fog
move	job	bomb	lose	shoes	shell	shock	moth	cough	soon
List 11	List 12	List 13	List 14	List 15					
badge	fish	fib	fill	kiss					
hutch	duck	thatch	catch	buzz					
kill	path	sum	thumb	hash					
thighs	cheese	heel	heap	thieve					
wave	race	wide	wise	gate					
reap	hive	rake	rave	wife					
foam	bone	goes	got	pole					
goose	wedge	shop	shown	wretch					
not	log	vet	bed	dodge					
shed	tomb	june	juice	moon					