Comparison of Different Hearing Aid Prescriptions for Children

Josephine E. Marriage[1], Deborah A. Vickers[2], Thomas Baer[3], Brian R. Glasberg[3], and Brian C.J. Moore[3]

[1]Chear, 30 Fowlmere Road, Shepreth, Royston, Herts SG8 6QS, UK

[2]University College London, Speech, Hearing and Phonetic Sciences, Wakefield Street, London. WC1N 1PF, UK

[3]Dept of Experimental Psychology, University of Cambridge, Downing St, Cambridge CB2 3EB, UK

Send correspondence to:

Brian C.J. Moore
email: bcjm@cam.ac.uk

Objective: To assess whether there are significant differences between speech scores for different hearing-aid prescription methods, specifically DSL i/o, DSL V, and NAL-NL1, using age-appropriate closed-set and open-set speech tests with young children, designed to avoid floor and ceiling effects.

Design: Participants were 44 children with moderate or severe bilateral hearing loss, eight aged 2-3 years, 15 aged 4-5 years, and 21 aged 6-9 years. Children wore bilateral hearing aids fitted with each prescription method in turn in a balanced double-blind design. The speech tests used with each child (and for some tests the levels) were chosen so as to avoid floor and ceiling effects. For the closed-set tests, the level used was selected for each child based on their hearing loss. The tests used were: (1) The closed-set consonant confusion test (CCT) of word identification; (2) The closed-set Chear Auditory Perception Test (CAPT) of word identification. This has separate sections assessing discrimination of consonants and vowels and detection of consonants; (3) The open-set Cambridge Auditory Word Lists (CAWL) for testing word identification at levels of 50 and 65 dBA, utilising 10 consonant-vowel-consonant real words that are likely to be familiar to children aged 3 years or older; (4) The open-set common phrases test (CPhT) to measure speech reception threshold (SRT) in quiet; (5) Measurement of the levels required for identification of the Ling 5 sounds, using a recording of the sounds made at the University of Western Ontario.

Results: Scores for CCT and CAPT consonant discrimination and consonant detection were lower for the NAL-NL1 prescription than for the DSL prescriptions. Scores for the CAPT vowel-in-noise discrimination test were higher for DSL V than for either of the other prescriptions. Scores for the CAWL did not differ across prescriptions for the level of 65 dBA, but were lower for the NAL-NL1 prescription than for either of the DSL prescriptions for the level of 50 dBA. The SRT measured using the CPhT and the levels required for identification of the Ling 5 sounds were higher (worse) for the NAL-NL1 prescription than for the DSL prescriptions.

Conclusions: The higher gains prescribed by the DSL i/o and DSL V prescription methods relative to NAL-NL1 led to significantly better detection and discrimination of low-level speech sounds.

INTRODUCTION

There is general agreement that the selection of frequency- and level-dependent gains for a hearing aid fitting for a child should be based on a prescription formula (Mueller et al. 1992). This paper presents a comparison of three hearing-aid fitting methods that were in widespread use with both adults and children at the time this study was conducted. These methods are the two versions of DSL (DSL i/o and the updated version V, also called DSL m[i/o]) (Cornelisse et al. 1995; Scollie et al. 2005) and NAL-NL1 (Byrne et al. 2001).

There are differences between the rationales and amplification characteristics of the NAL-NL1 and the DSL prescription methods: NAL-NL1 generally prescribes less low- and high-frequency gain than the DSL methods, particularly for severe or profound hearing loss, as shown in Table 1.

TABLE 1. Comparison of real-ear aided response (REAR) targets prescribed by DSL V, DSL i/o and NAL-NL1 for a child with a flat moderate loss (left) and a sloping severe loss (right) for input levels of 50, 65 and 75 dB SPL

Flat moderate loss							Sloping severe loss						
DSL V	Frequency, Hz							Frequency, Hz					
Level	250	500	1000	2000	4000	6000	Level	250	500	1000	2000	4000	6000
50	71	72	69	74	74	71	50	62	71	76	86	85	85
65	79	82	78	84	84	82	65	72	80	88	101	100	100
75	80	85	86	93	92	87	75	73	84	94	108	107	104
DSL i/o													
Level	250	500	1000	2000	4000	6000	Level	250	500	1000	2000	4000	6000
50	72	78	74	74	77	75	50	63	78	84	93	87	87
65	80	85	81	84	87	81	65	73	83	94	105	102	101
75	80	87	87	90	93	87	75	73	86	92	109	108	108
NAL-NL1													
Level	250	500	1000	2000	4000	6000	Level	250	500	1000	2000	4000	6000
50	50	64	70	73	68	63	50	48	65	76	79		
65	63	73	77	82	78	71	65	58	73	83	91		
75	71	78	82	87	86	81	75	66	79	88	99	89	

Empty cells indicate that the prescription formula did not give a target value.

NAL-NL1 also prescribes less compression than DSL, especially at high frequencies. The amount of compression in hearing aids represents a compromise; for reviews, see Dillon (1996) and Moore (2007; 2008). Compression is required to ensure that low-level sounds are audible while intense sounds are not uncomfortably loud, and this sometimes requires high compression ratios. Greater compression is required to give access to speech cues for people with severe and profound hearing loss, who often have a very small dynamic range between hearing thresholds and discomfort levels. However, high compression ratios can have deleterious effects, especially when fast-acting compression is used. Such effects include reduced amplitude modulation depth (Plomp 1988; Stone & Moore 1992), reduced spectral contrast (Plomp 1988), and “cross-modulation” between different sound sources (Stone & Moore 2004). High compression ratios, combined with high amounts of low-frequency gain, may also increase the audibility of background noise, and this may degrade speech understanding in noise via the upward spread of masking. Thus, as compression ratios are increased, the potential benefits of increased audibility of speech may be offset by a variety of deleterious effects. Compression at high frequencies has been found to be beneficial for some hearing-impaired adults, but the results for children are less clear cut (Marriage & Moore 2003; Marriage et al. 2005). The optimal amount of compression for children remains unclear.

There are several methods that could in principle be used for comparing the effectiveness of different hearing-aid prescription measures. One method is based on the Aided Speech Intelligibility Index (SII) (ANSI 1997; Stelmachowicz et al. 2000; Ching et al. 2001; McCreery & Stelmachowicz 2011; Stiles et al. 2012a), which is a measure of the audibility of speech when amplification is provided by a hearing aid. Davidson and Skinner (2006) reported that SII values were correlated with aided speech intelligibility for school-age children. Stiles et al. (2012a) reported that, for children aged 6-9 years, the Aided SII was a better predictor of word and non-word repetition and receptive vocabulary than the pure-tone average across the frequencies 0.5, 1, 2, and 4 kHz. However, greater stimulus bandwidths and higher sensation levels (i.e. higher SII values) are necessary for children to achieve similar performance to adults (Stelmachowicz et al. 2001; Scollie 2008; McCreery &

Stelmachowicz 2011). One problem with this approach is that accurate estimates of audiometric thresholds, which are required for calculation of the Aided SII, may not be available for young children. Also, the Aided SII does not take into account the effect of supra-threshold discrimination problems, such as reduced frequency selectivity and impaired temporal processing (Moore 2007; 2014), that can have a strong influence on the ability to understand speech. Finally, it is difficult to calculate the SII when the hearing aids incorporate nonlinear processing such as fast-acting amplitude compression or frequency lowering.

A second approach for comparing hearing-aid prescription methods is via questionnaire measures of functional auditory skill development, for example the Meaningful Auditory Integration Scale (MAIS) (McConkey Robbins et al. 2004), the Auditory Skills Checklist (ASC) (Meinzen-Derr et al. 2007), Parents’ Evaluation of Aural/oral performance of CHildren (PEACH) (Ching & Hill 2007), the Self Evaluation of Listening Function (SELF) (Ching et al. 2008), and the University of Western Ontario Pediatric Audiological Monitoring Protocol (UWO PedAMP) (Bagatto et al. 2010). These are used to quantify auditory and vocal behaviors. Using this approach, Ching et al. (2010b) found that NAL-NL1 led to better rated performance for speech in noise than DSL v.4.1 for both PEACH and SELF, although the difference was significant only for SELF for children tested in Australia. Ching et al. (2013) compared two groups of children, one fitted with NAL-NL1 and the other fitted with DSL v.4.1. Questionnaire measures of vocabulary and expressive and receptive language did not differ significantly across the two groups. A problem with the use of questionnaires is that the outcomes may be influenced by the personality and attitude of the adult or child performing the evaluation. Hence, questionnaires may be useful for comparing results across groups, but are not so effective in evaluating the performance of individual children. Also, little insight is gained into the supra-threshold auditory processing abilities of the child.

A third approach to comparing hearing-aid prescription methods is via the use of paired comparisons of the intelligibility of speech (Ching et al. 2010b; Moore et al. 2011; Moore & Sek 2013). Ching et al. (2010b) used this approach to compare NAL-NL1 to DSL v.4.1 fittings for children tested in Australia and Canada. Of the children tested in Australia, 17 out of 24 showed a significant preference for one fitting over the other (10 for NAL-NL1 and 7 for DSL v.4.1). Of the children tested in Canada, 16 out of 24 showed a significant preference (8 for NAL-NL1 and 8 for DSL v.4.1). A limitation of this approach is that preferences may be influenced by whatever prescription each child had been using most recently.

A more direct approach to comparing hearing-aid fitting procedures is via the use of tests of the ability to discriminate and understand speech. A study comparing DSL 4.1 and NAL-NL1 prescriptions for older children (6.6 to 19.8 years old), including measures of the ability to understand consonants in quiet and sentences in noise, showed no clear overall benefit for one prescription over the other (Ching et al. 2010a). However, a more recent study using children with more severe hearing loss, showed better performance with, and preferences for, DSL V over NAL-NL1 for children aged from seven to 17 years (Quar et al. 2013). We are not aware of any previous comparisons of NAL-NL1 and DSL using speech testing for children aged 6 years or below.

A problem in assessing speech perception abilities for young children is the selection of age-appropriate tests so as to avoid floor and ceiling effects (Govaerts et al. 2006). Many studies on early intervention for hearing-impaired children do not report speech recognition scores for participants younger than about 6 years, demonstrating either limited perceived validity or difficulties in acquiring the data (Strauss & van Dijk 2008). The use of open-set speech tests requires sufficiently clear articulation by the hearing-impaired child to allow valid and reliable scoring (Stiles et al. 2012b). Basic phonological reading skills are required for written response options in nonsense word tests (Scollie 2008). Both of these methods of speech testing place a lower limit on the age at which valid and repeatable testing is possible. Additional constraints arise from the short attention span of young children and the related difficulty in maintaining interest and therefore compliance with the task. Tests need to have a sufficient number of items to give small critical differences between test scores (Thornton & Raffin 1978; Vickers et al. 2018) and to give good test/retest reliability (Bland & Altman 1986; Lovett et al. 2013), but they should not be so long that the child loses interest and/or concentration. As a result of these problems "few clinically useful measures exist to evaluate auditory development in infants and toddlers", regardless of hearing status (McConkey Robbins et al. 2004).

In a companion paper (Vickers et al. 2018) we describe two speech tests that can be used with children aged from 2 to 9 years so as to obtain meaningful results and to avoid floor and ceiling effects. These tests, together with some others, were used in the present study to compare the effectiveness of DSL i/o, DSL V and NAL-NL1 for children from two to nine years of age.

METHOD

This research was approved by the Cambridge Research Ethics Committee (Approval number 06/Q0108/321).

Speech Test Selection

The monosyllabic closed-set speech tests are described in detail in our companion paper (Vickers et al. 2018), so only a brief description is given here.
The tests are summarised in Table 2 and were:
(1) The consonant confusion test (CCT). This is a closed-set monosyllabic word test made up of sets of four words, represented by pictures, which are familiar to children from about 2 years of age.
(2) Three closed-set monosyllabic word tests with pictures called the Chear Auditory Perception Test or CAPT. The CAPT has separate sections assessing discrimination of consonants and vowels and detection of consonants. All closed-set tests (i.e. CCT and CAPT) had four response options, depicted by a picture with the target word written underneath. The pictures were presented on a touch-sensitive screen or on a laptop with a mouse and the child was asked to select the picture corresponding to the word that was heard through the loudspeaker.
(3) Ten open-set consonant-vowel-consonant (CVC) real word lists for use in quiet and/or noise (called the Cambridge Auditory Word Lists or CAWL), each using 10 words that

would be familiar to children aged 4 years or more. The words in each list were phonetically balanced and one list was presented for each condition (each prescription method). For open-set testing, children spoke their responses, which were written down by the tester and scored. Open-set testing was video recorded for later review, if needed, although this was not typically required for accurate scoring. Tests using normal-hearing children aged from 3 years 11 months to 8 years 3 months showed that, for a level of 50 dBA (the lower of the two levels used in this study), phoneme scores ranged from 24 to 30 out of 30 (unpublished data). The score for each normal-hearing child was converted to RAU (Studebaker 1985), and the mean and standard deviation of the transformed scores was calculated and then transformed back to scores out of 30. The mean was 29 out of 30 with a standard deviation of 3.

(4) The common phrases test (CPhT) recorded using a UK English speaker with a British English accent and typical English vocabulary (Robbins et al. 1988). This test was used to estimate the speech reception threshold (SRT) in quiet.

(5) Measurement of the levels required for identification of the Ling 5 sounds (the speech sounds /u ɑ i ʃ s/) (Scollie et al. 2012; Glista et al. 2014).

The Renfrew word finding test (Renfrew 1995) was used as a vocabulary screen to determine the parts of the speech test battery to use with each child and to maintain an appropriate level of challenge and thereby self-motivation for each child. With this test, the child is asked to name the picture presented on each of 50 cards. Children in the age range 2-3 years were all tested using the CCT because the pictorial representations do not require reading skills for their identification. The maximum score is 50. Normative data are given in Renfrew (1995). Older children with Renfrew scores in the range 9-25 were also tested using the CCT, as were children who achieved scores of 26-36 but who could not read. Children with early grapheme recognition and vocabulary scores of over 26 were tested using the CAPT. We were not always successful in obtaining useful results for all of the tests that were initially selected for each child. However, data are reported for all tests that were completed for all three prescription methods for a given child.

TABLE 2. The number of children tested with each speech perception test, separated by age group and severity of loss

Speech test	Speech material	Age: 2 – 3 years (n=8)	Age: 4 – 5 years (n=15)	Age: 6 – 9 years (n= 21)
CCT	cow, owl, house, mouse; bed, hen, peg, egg; fan, man, cat, hat; key, three, feet, sheep; pig, chick, fish, ship; horse, ball, fork, door; shoe, moon, spoon, food; pipe, pie, kite, five; sock, cot, doll, dog; jug, duck, bus, cup	Mod n=4 Sev n=4	Mod n=2 Sev n=4	
CAPT Disc. subtest	mat, bat, cat, fat; wine, wise, white, wipe; fin, tin, shin, chin; stork, talk, chalk, fork; bun, bug, bud, buzz; kick, tick, thick, pick; white, right, light, night;		Mod n=5 Sev n=2	Mod n=15 Sev n=5
CAPT Vowel subtest	two, tea, tie, tar; beak, buck, bark, book; cart, cat, cut, cot;		Mod n=5 Sev n=4	Mod n=15 Sev n=5
CAPT Detection subtest	bee, bean, bees, beef slice, ice, lice, eye suit, shoot, shoe, sue		Mod n=8 Sev n=5	Mod n=15 Sev n=5
CAWL 50 dBA	10 lists of 10 open-set monosyllabic real words, scored by number of phonemes correct out of 30	Mod n=2 Sev n=0	Mod n=8 Sev n=3	Mod n=16 Sev n=4
CAWL 65 dBA		Mod n=2 Sev n=1	Mod n=8 Sev n=4	Mod n=16 Sev n=4
CPhT	Common phrases marked out of three key words, adaptive presentation level using 2-dB step size	Mod n=1 Sev n=0	Mod n=8 Sev n=5	Mod n=16 Sev n=5
Ling sounds level for identification	/u ɑ i ʃ s/	Mod n=4 Sev n=4	Mod n=8 Sev n=7	Mod n=16 Sev n=5

Mod and Sev indicate moderate and severe hearing loss, respectively. CCT: consonant confusion test. CAPT: Chear Auditory Perception Test. CAWL: Cambridge Auditory Word Lists. CphT. Common phrases test.

Speech Test Presentation

Speech materials were presented from a HP Compaq nx7400 laptop computer via an Edirol UA-1ex USB Audio interface sound card. This fed into a Kamplex AC35 2-channel audiometer, so that output levels could be adjusted through each channel in 1-dB steps. The audiometer output was fed to a Mordaunt-Short 902 loudspeaker placed at about 1 meter distance in front of the child, in audiology test rooms fulfilling ISO 8253-3 in six different clinical venues in the UK. Calibration of levels was carried out using a stored noise file with the same average spectrum and level as the CAPT test items. The audiometer VU meter was set to 0 dB, the noise file was played and the sound level was measured using a sound level meter close to the listening position of the child.

Measures Obtained

The measures obtained were:

(1) Percent correct and discriminability index (d′) scores (Macmillan & Creelman 2005) for the closed-set CCT. The value of d′ increases monotonically with percent correct for a given number of response alternatives, and it increases monotonically with the number of alternatives for a fixed percent correct. The value of d′ can be readily obtained from standard tables (Hacker & Ratcliff 1979). There are two advantages of d′ over percent correct: d′ scores are less affected than percent correct scores by floor and ceiling effects; and d′ scores can be meaningfully compared across tests with different numbers of response alternatives.

(2) Percent correct and d′ scores for the closed-set CAPT. The vowel sub-test was conducted in the presence of a speech-shaped background noise with a speech-to-noise ratio of 0 dB, to increase the difficulty of the test, thus avoiding ceiling effects. For each sub-test there were six different orders of the groups of words, one of which was randomly selected for each condition for each child.

3) For the CCT and CAPT tests, the percentage transmission of voicing, place and manner information for consonants was calculated from confusion matrices for each word group. It should be borne in mind that, because there were only four response alternatives for each

consonant, not all possible confusions were allowed for. Hence, the analyses only represent confusions within the limited number of alternatives available. Also, for the CCT, the response alternatives differed from the target in both their initial and final consonants, and a misperception of one of these might influence the decision about the other. Hence, the consonant confusions need to be interpreted with caution. Vowel confusions were analysed to estimate errors in the features of height, place and duration. The values were averaged for each feature across the test list and scores were converted to d′ values.

(4) Scores out of 30 for the open-set CAWL were converted to percent correct. They were not converted to d′ values, since d′ is not well defined for an open-set test. The CAWL words were presented in quiet at 50 and 65 dBA. A list of 10 familiar CVC monosyllabic words was presented and the child was asked to repeat each word. Responses were marked for the number of phonemes correct out of three for each word. Each list was only used once, so test items were novel for each hearing aid prescription. There were twelve lists, and the lists were used in a counter-balanced order across children.

(5) SRT in quiet for the CPhT (Robbins et al. 1988). This test was replayed from a compact disc produced by the cochlear-implant team at St Thomas's Hospital, London. After each phrase was presented, the child was asked to repeat it. Scoring was by number of key words correct, with three key words per phrase. An adaptive paradigm was used to adjust the presentation level so as to estimate the SRT. The level was decreased when two or three words were correctly identified and increased when either one or no words were correctly identified. The step size was 5 dB until two turnpoints had occurred and was 2 dB thereafter. At least two turnpoints with the 2-dB step size were obtained, and the SRT was defined as the mean level over the final two turnpoints.

(6) Minimal levels of presentation required for correct repetition of the pre-recorded Ling sounds /u a i ʃ s/. Each of the sounds was presented in isolation and the child was required to repeat the sound. The sequence of presentation of the different sounds was randomized, and the levels were also randomized over a range from below the detection threshold to well above it. For each sound, the lowest level at which the sound was correctly repeated was determined. The sixth Ling sound /m/ was omitted from the test sounds, as /m/

could not be reliably discriminated from /u/ by normal-hearing control children with ages 2-3 years.

Individual children were able to complete different subsets of the speech tests, depending on their hearing, clarity of articulation, and attention skills, particularly for the open-set speech tests. Only some of the children under 4 years of age were tested with the open-set tests.

All of the closed-set word tests have long and short versions. The short forms of both the CAPT and the CCT contain 40 words intended to be appropriate for children with developmental ages of three years or more. The longer forms contains 32 additional words that are appropriate for children with developmental ages of five years and above. In the analyses presented below, only data for the short versions of the tests were used, since these were completed by greater numbers of children.

Speech Presentation Levels for Closed-Set Tests

For the closed-set materials, the performance-intensity function is very steep (steeper than for the open-set CAWL test), and the level leading to any specific performance varies markedly across children. This makes it impractical to use a fixed testing level for all children, since some would perform close to chance and others would perform close to ceiling. Hence, for all of the closed-set tests except CAPT vowel discrimination in noise, fixed presentation levels were used for each child, but the level used varied across children. It is not feasible to restore the audibility of low-level sounds completely to normal for hearing-impaired children or adults, due to factors such as the internal noise of hearing aids (especially microphone noise), limitations in the gain that can be achieved without acoustic feedback, and the need to avoid excessive amounts of compression. In practice, the lowest sound level for which audibility can be restored needs to increase with increasing hearing loss (Keidser et al. 2011). We wished to avoid floor or ceiling effects for the closed-set speech materials, by presenting the stimuli at a relatively low level, but not so low that limited audibility would severely compromise performance. To achieve this, the presentation level in dBA of the closed-set test material was chosen for each child based on their unaided hearing levels, using the following formula:

$$\text{Level} = [\text{PTA in better hearing ear} \times 0.4] + 30$$

where PTA (pure-tone average) is the mean audiometric threshold for the three worst thresholds out of 500, 1000, 2000 and 4000 Hz for the better-hearing ear. The constant “30” was based on our finding that the lowest level at which children with normal hearing could complete the closed-set tests in quiet was about 30 dBA. The slope of “0.4” was chosen so that, with increasing PTA, the presentation level would increase more slowly than the PTA (Keidser et al. 2011). For example, if the audiometric thresholds at 500, 1000, 2000 and 4000 Hz were 30, 35, 40, and 45 dB HL, the PTA was taken as 40 dB HL, and the presentation level was $(40 \times 0.4) + 30 = 46$ dBA. If the PTA was 60 dB HL, the presentation level was (60 x 0.4) + 30 = 54 dBA.

One of the closed-set tests, CAPT vowel discrimination, was performed using a fixed level of 60 dBA, and the vowels were presented in a speech-shaped background noise at 0 dB signal-to-noise ratio. It was judged that, for this test and for the noise level used, performance would be mainly determined by the signal-to-noise ratio, rather than by the absolute level.

Children and Test Conditions

Fifty-four children were initially enrolled into the study and 44 children completed speech testing for each prescription condition for their age group. Of the ten children who dropped out of the study, only one dropped out due to an inability to complete the speech testing. This child had more global communication difficulties. The main reason for dropping out of the study was family difficulties in attending the required five appointments. The hearing losses of the 44 remaining children were classified as moderate or severe based on the average audiometric thresholds for the better ear over the frequencies 500, 1000, 2000 and 4000 Hz. Those with an average in the range 35 to 65 dB HL were classified as moderate and those with an average in the range 66 to 95 dB HL were classified as severe. The children were divided into three age groups:

Group 1 (2-3 yrs): n=8 (4 moderate, 4 severe)

Group 2 (4-5 yrs): n=15 (8 moderate, 7 severe)

Group 3 (6-9 yrs): n= 21 (16 moderate, 5 severe)

The range of ages was from 2 years 7 months to 9 years 8 months. The highest age was originally intended to be eight years, but one 9-year-old asked to be included as both of her brothers were enrolled in the study. All others were 8 years or under.

All children were initially tested on all of the tests using their own hearing aids, as fitted by their own audiologist. Most of the children had been fitted using DSL i/o targets, following the “Modernisation of children's hearing aids” protocol that was widely adopted in the UK, but a few may have been fitted with NAL-NL1; we did not have definite information about the previous fitting for some of the children. However, our measures of hearing-aid gain showed that their own aids often did not match targets for either DSL i/o or NAL-NL1. Generally, the measured gains were below the target gains, especially when compared to DSL i/o targets. Also there was a lot of variability across children in the deviation of the fittings from DSL i/o or NAL-NL1 targets. Hence, the initial testing with their own aids was considered as practice in performing the speech tests, and the data are not presented.

Several hearing aid types were used for the study. Hearing aids were chosen to be compatible with the wireless system for reception of the teacher’s voice that each child was using in school at the time of enrolment into the study. The hearing aids used were: Savia Art, Eterna and Naida aids manufactured by Phonak and Safran and Spirit P aids manufactured by Oticon. New closely fitting earmolds were made for each child so as to minimize acoustic feedback. This was important to allow the implementation of prescriptions with more high-frequency gain, which otherwise might have led to acoustic feedback. These earmolds were used throughout the study. Real-ear-to-coupler difference (RECD) measurements were made for each child to incorporate the acoustic effects of the earmold. If a new ear mould was required over the course of the study, the RECD was re-measured and incorporated into the prescription fitting. All hearing thresholds were measured using inserts attached to each child's own molds.

The hearing aid gains were adjusted to match targets for NAL-NL1, DSL i/o and DSL V, the gains for each prescription being stored in the hearing aid programming software under a blind code. The targets for the hearing aid prescriptions were derived through

Audioscan Verifit or Audioscan RM500 electroacoustic analysers according to ANSI S3.22 (1996) and published hearing aid fitting procedures (Cornelisse et al. 1995; Byrne et al. 2001; Scollie et al. 2005). The input signal for the Verifit was real speech, the “carrot passage” presented at 65 dB SPL. The hearing aid output was recorded through a 2-cc coupler. Hearing aid options such as directional microphone, noise reduction, frequency compression and feedback cancellation were deactivated, unless they were activated in the hearing aids that the child wore just before taking part in the study; this was a recommendation of the Ethical Board that approved the study. Only four children used hearing aids with frequency compression. Targets were matched as closely as possible (within ± 3 dB) across the octave frequency bands. It was nearly always possible to achieve a match within ± 3 dB for the center frequencies of 500, 1000 and 2000 Hz. For the center frequency of 250 Hz, most matches were within ± 3 dB, but a few fell outside that range, from –5 to +6 dB. For the center frequency of 4000 Hz, about 2/3 of cases fell within ± 3 dB, the remainder falling in the range –22 to +6 dB. Gains that were 10 dB or more below targets occurred when the hearing loss was 85 dB or more at 4000 Hz. The recommended maximum power output was matched as closely as possible for an 80-dB SPL swept tone, within the constraints of the hearing aid output. When NAL-NL1 did not prescribe a high-frequency target for a given frequency band, the hearing aid gain was not adjusted from the manufacturer’s pre-set value for that frequency band.

The programs were coded as C1, C2 and C3, with a random assignment of fitting method to program number. The tester was blind as to which program number corresponded to a given prescription. The order of activation and testing with the different prescriptions was randomised across children to control for order effects. Each prescription was tested roughly equally often in first, second, and third order.

The time allowed for each child to become familiar with the amplification characteristics for a specific prescription was selected bearing in mind the time available for the study and the changing listening skills of the child with increasing age. Children wore the study hearing aids with each prescription in turn, typically for between 2 and 4 weeks for each prescription. At the end of this familiarization/acclimatization period, they were assessed using the speech test battery and the next prescription was programmed in. The tester was blind to the prescription being used at the time of testing and when programming the next prescription condition.

RESULTS

For all of the speech tests, the results were analysed using repeated-measures analysis of variance (ANOVA) with factor prescription formula (DSL i/o, DSLV and NAL-NL1). The dependent variables in the ANOVAs were the d′ values for manner, place, voicing and overall score (except for vowels, where voicing was replaced by vowel height and manner was replaced by duration), identification level, or SRT. For the CAWL, the percent correct scores were converted to rationalized arcsine units (RAU) for analysis (Studebaker 1985). Mauchley's test of sphericity was applied. When the condition of sphericity was not satisfied, the Greenhouse-Geisser correction was used to modify the degrees of freedom. Pairwise comparisons were conducted using Fisher's protected least significant difference test (with $p<0.05$ as the criterion level). The effect size has been calculated for each analysis where the effect was significant and the eta squared (η^2) value is reported. These values are interpreted using the classification from Cohen (1988) for ANOVA, whereby the small effect size boundary falls at 0.01, the medium effect size boundary is at 0.06 and the large effect size boundary is at 0.14. For the CAWL, for which the same two levels were used for all children, the results for each level were analyzed with severity of hearing loss as a between-subjects factor. For the other tests, for which levels were chosen separately for each child (the level increasing with increasing hearing loss), the results were not separated according to severity of hearing loss.

Results for Closed-Set Speech Tests

The results for all of the closed-set speech tests are summarized in Table 3.

TABLE 3. Mean percent correct and d′ scores with standard errors for all of the closed-set tests: CCT and CAPT, separated into the discrimination, vowel and detection categories

Test	Score type	DSL i/o	DSL V	NAL-NL1	F-value and significance
CCT n=14,	% correct	82.9	80.4	74.1	F(2,26)=5.81
Age=2:7 - 4:10	d′	2.1 (0.1)	2.0 (0.1)	1.6 (0.1)	p<0.01
CAPT Disc n=27	% correct	80.7	77.5	71.0	F(2,54)=9.27
Age=5:4 - 9:8	d′	2.0 (0.1)	1.9 (0.1)	1.6 (0.1)	p<0.01
CAPT Vowel n=23	% correct	87.7	92.8	86.2	F(2,44)=2.58
Age=5:4 – 9:8	d′	2.2 (0.2)	2.6 (0.2)	2.2 (0.2)	p=0.09
CAPT Det n=28	% correct	88.2	81.3	76.2	F(2,54)=9.33
Age=2:11 – 9:8	d′	2.8 (0.2)	2.4 (0.2)	2.0 (0.1)	p<0.001

Outcomes of ANOVAs based on the d′ scores with "prescription" as the single factor are given in the right-most column

Consonant Confusion Test (CCT) for youngest children • Fourteen children aged 2-4 years were tested with the CCT. The mean percent correct scores are shown on the right axis of Figure 1, with corresponding d′ values on the left axis. The dashed lines represent scores of 25, 50, 75, and 100%. There was a significant effect of prescription formula for consonant place of articulation ($F(2,26) = 6.57$, $p<0.01$; $\eta^2 = 0.15$ (large effect size) and overall ($F(2,26) = 5.81$, $p<0.01$; $\eta^2 = 0.17$ (large effect size). Pairwise comparisons showed that the DSL prescriptions both gave significantly higher scores than the NAL-NL1 prescription, but scores for the two DSL prescriptions were not significantly different from one another. There was no significant effect of prescription formula for manner ($F(2,26) = 2.91$, $p = 0.07$) or voicing ($F(2,26) = 2.44$, $p = 0.11$).

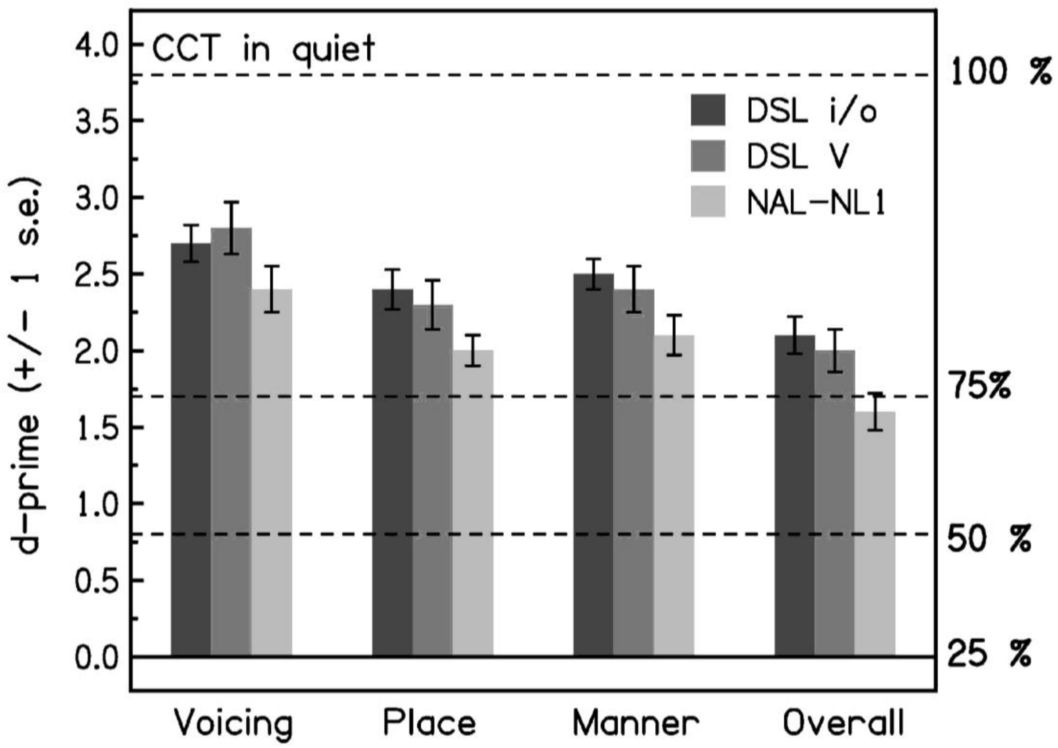

Figure 1. Mean percent correct scores (right-hand axis) and d′ values (left-hand axis) for voicing, place, manner and overall for the CCT. In this and all subsequent figures, error bars indicate ±1 standard error. In this and similar later figures, the horizontal lines represent scores of 25, 50, 75, and 100%.

CAPT Consonant discrimination • Figure 2 shows mean percent correct scores and d′ values for voicing, place, manner and overall for CAPT discrimination of consonants. There were significant effects of prescription for manner ($F(2,54) = 8.48$, $p<0.01$; $\eta^2 = 0.06$ (medium effect size), place, ($F(2,54) = 4.64$, $p<0.01$; $\eta^2 = 0.05$ (small effect size) and overall ($F(2,54) = 9.27$, $p<0.01$; $\eta^2 = 0.08$ (medium effect size), but not for voicing. Pairwise comparisons for manner and overall score showed that scores for the two DSL prescriptions were significantly higher than for the NAL-NL1 prescription, but were not significantly different from one another. Pairwise comparisons for place showed that the mean score for DSL i/o was significantly higher than for NAL-NL1 but not than for DSL V, and that scores for DSL V and NAL-NL1 were not significantly different from one another.

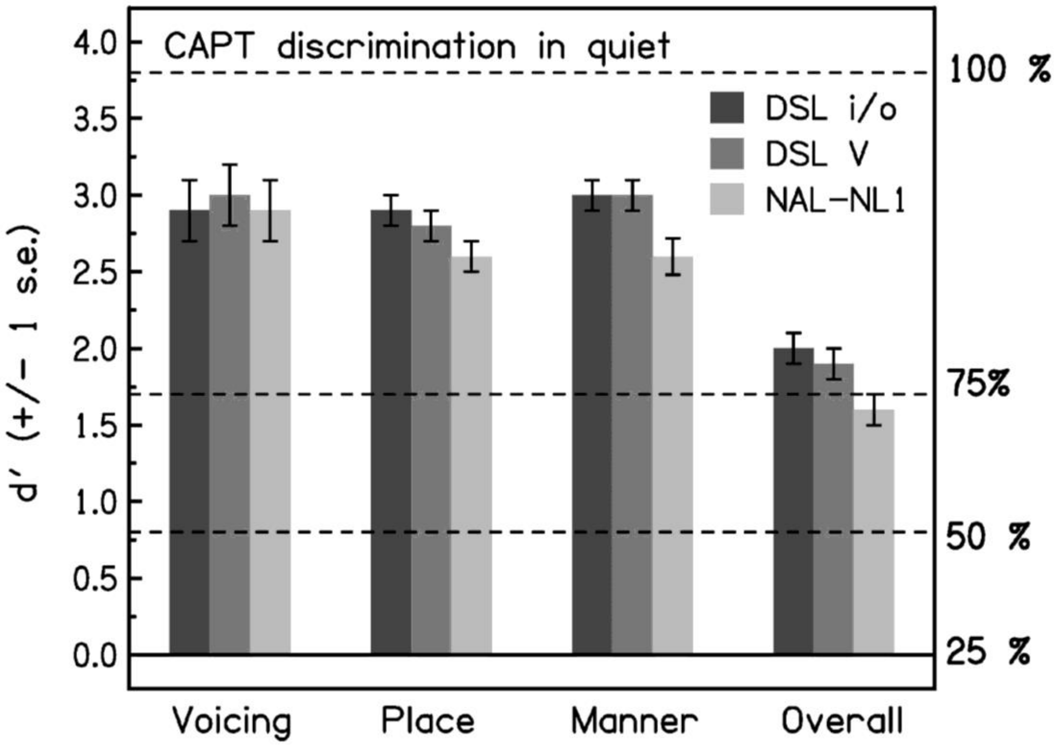

Figure 2. Mean percent correct scores (right-hand axis) and d′ values (left-hand axis) for voicing, place, manner and overall for CAPT consonant discrimination.

CAPT Vowel Discrimination in Noise • For the analysis of errors within the vowel sets, "height" relates to the openness of the articulation (which partly determines the frequency of the second formant), "place" refers to whether the main point of narrowing in the vocal tract was front, mid or back (which partly determines the frequency of the first formant), and "duration" refers to whether the vowel was short, long or a diphthong. Twenty three children completed assessments with the vowel test materials in noise.

The mean scores are shown in Figure 3. There were significant effects of prescription for height ($F(2,44) = 3.33$, $p<0.05$; $\eta^2 = 0.04$ (small effect size) and place ($F(2,44) = 4.38$, $p = 0.03$; $\eta^2 = 0.04$ (small effect size), but not for duration ($F(2,44) = 2.85$, $p = 0.07$) or overall ($F(2,44) = 2.58$, $p = 0.09$). Pairwise comparisons for height showed that the mean score for DSL V was significantly higher than for NAL-NL1 but not than for DSL i/o. Scores did not differ significantly between NAL-NL1 and DSL i/o.

Pairwise comparisons for place showed that the mean score for DSL V was significantly higher than for both DSL i/o and NAL-NL1, but that scores for DSL i/o and NAL-NL1 were not significantly different from one another.

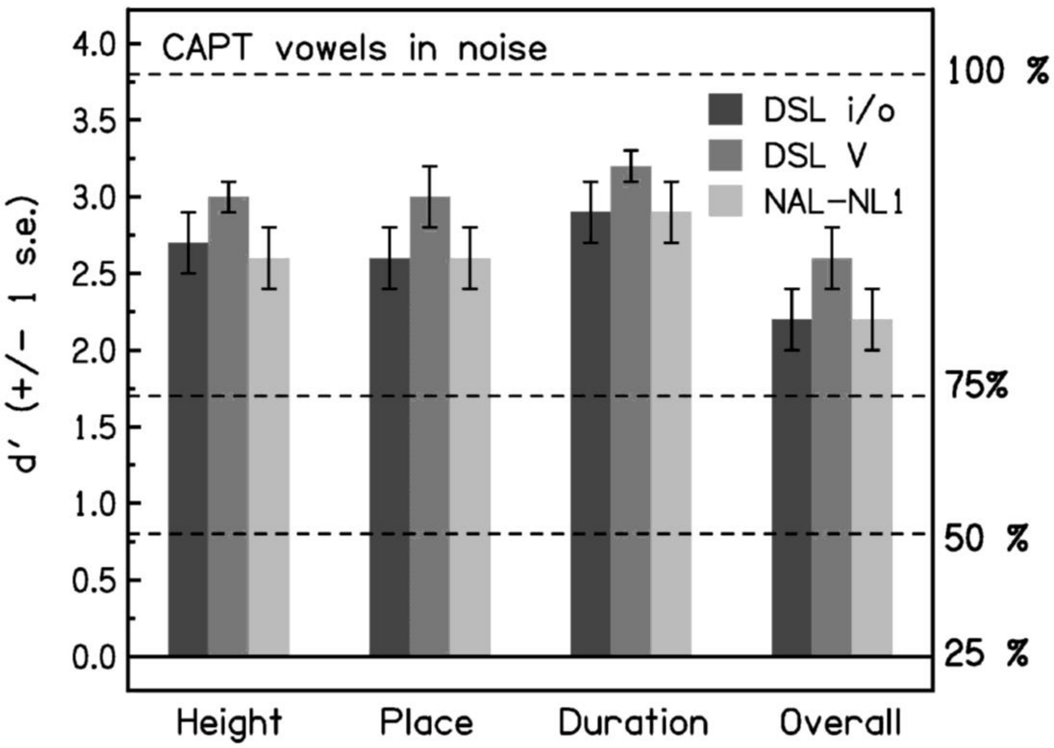

Figure 3. Mean percent correct scores (right-hand axis) and d′ values (left-hand axis) for height, place, duration and overall for CAPT vowels in noise.

CAPT Detection of Consonants • Twenty-eight children took the CAPT detection test. Scores for "manner" were based both on the standard categories of manner of articulation (plosive, fricative, approximant, etc.) and on whether a speech sound was present or not, e.g. eye compared to ice. The mean scores are shown in Figure 4.

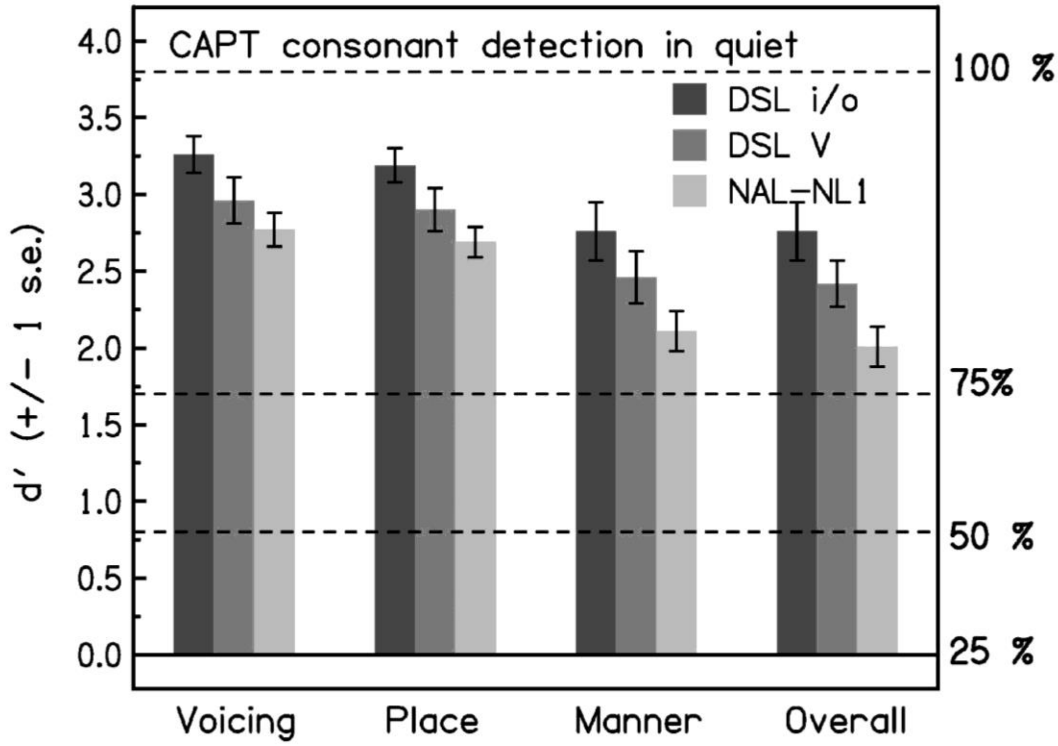

Figure 4. Mean percent correct scores (right-hand axis) and d′ values (left-hand axis) for voicing, place, manner and overall for CAPT detection.

There were significant effects of prescription formula for voicing ($F(2,54) = 7.54$, $p<0.01$; $\eta^2 = 0.08$ (medium effect size), place ($F(2,54) = 7.70$, $p<0.01$; $\eta^2 = 0.10$ (medium effect size), manner ($F(2,54) = 6.78$, $p<0.01$; $\eta^2=0.10$ (medium effect size), and overall ($F(2,54) = 9.33$, $p<0.001$; $\eta^2 = 0.14$ (large effect size). Pairwise comparisons for voicing, manner and overall showed that DSL i/o and DSL V gave significantly higher scores than NAL-NL1, but scores for DSL i/o and DSL V were not significantly different from one another. For place, the score was significantly higher for DSL i/o than for NAL-NL1, but the score for DSL V did not differ from scores for DSL i/o or NAL-NL1.

Results for Open-Set Speech Tests

CAWL scores were derived from the number of phonemes correct for each of the target words. For the CPhT, each score is the SRT in dBA. The results are summarized in Table 4.

TABLE 4. Mean scores and standard errors for children tested with each open-set test

Test	Measurement units	DSL i/o	DSL V	NAL-NL1	F-value and significance
CAWL 50 dB n = 32 Age = 3:11 – 9:8	Score out of 30	23.2 (0.6)	22.8 (0.9)	19.7 (0.8)	F(2,62)=10.30 p<0.001
CAWL 65 dB n = 32 Age = 3:11 – 9:8	Score out of 30	27.0 (0.5)	27.0 (0.5)	26.9 (0.7)	F(1.7,51.7)=0.04 p=ns
CPhT n = 35 Age = 3:11 – 9:8	SRT in dBA	38.7 (1.2)	39.4 (1.3)	41.3 (1.22)	F(2,68)=7.1 p=0.002

Outcomes of ANOVAs based on the d′ scores for the factor "prescription" are given in the right-most column

CAWL • Unlike the closed-set tests, for which the stimulus level was chosen for each child according to the severity of that child's hearing loss, the stimuli for the CAWL were presented at the same level for all children. This made it meaningful to score the results separately for the two severities of hearing loss, moderate and severe. Figure 5 shows the mean score for each prescription for each severity group and each presentation level. A few children with moderate hearing loss scored close to ceiling for the 65-dB SPL stimuli. Otherwise, scores were below ceiling. ANOVAs were conducted separately on the RAU-transformed scores for the presentation levels of 50 and 65 dBA with prescription as a within-subjects factor and severity of hearing loss as a between-subjects factor. For the level of 65 dBA, there was no significant effect of prescription ($F(1.7, 51.7) = 0.04$, $p = 0.95$), but there was an effect of severity of hearing loss ($F(1, 30) = 12.44$, $p = 0.001$). Children with moderate hearing loss had higher scores than those with severe hearing loss. There was no significant interaction.

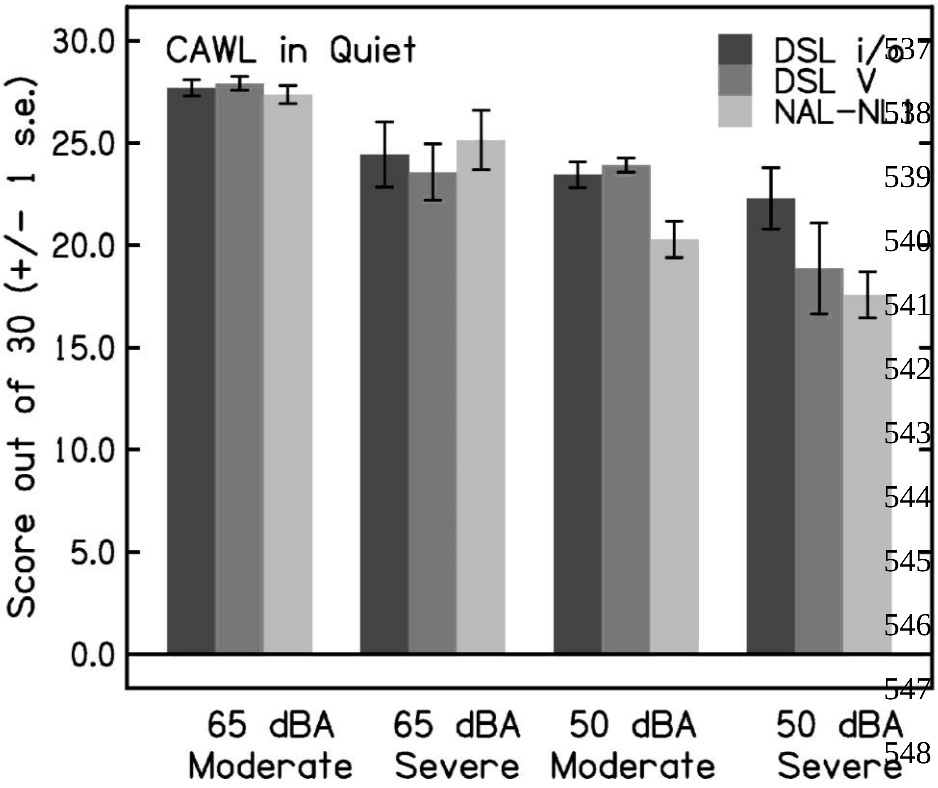

Figure 5. Mean score out of 30, with children divided into two groups according to severity of loss (moderate or severe) for the CAWL presented at 50 and 65 dBA.

For the level of 50 dBA, there was a significant effect of prescription ($F(2,62) = 10.30$ $p<0.001$) and a significant effect of severity of hearing loss ($F(1,30) = 5.2$, $p = 0.03$). There was no significant interaction. As for the 65-dBA level, children with moderate hearing loss had higher scores than those with severe hearing loss. Pairwise comparisons showed that scores for DSL i/o and DSL V were significantly higher than those for NAL-NL1, but were not significantly different from one another.

CPhT *• The mean SRTs for the CPhT are given in table 4. The ANOVA showed a significant effect of prescription formula (*$F(2,68) = 7.1$, $p<0.002$; $\eta^2 = 0.02$*, small effect size). Pairwise comparisons showed that DSL i/o and DSL V gave significantly better scores than NAL-NL1, but scores for DSL i/o and DSL V were not significantly different from one another.*

Level Required for Ling Sound Identification

Figure 6 shows the average levels in dBA required for correct identification of each of the Ling sounds, across all 44 subjects, for each hearing aid prescription. An ANOVA with prescription formula and sound as factors showed significant main effects of prescription ($F(1.5,65.3) = 9.66$, $p<0.001$) and sound ($F(2.5,106) = 61.72$, $p<0.001$). There was also a significant interaction between sound and prescription ($F(5.8,251) = 5.5$, $p<0.001$), confirming that the effect of prescription differed across Ling sounds. The vowel sounds /u/ and /i/ and the high-frequency fricative /s/ were identified at significantly lower levels with the DSL prescriptions than with the NAL-NL1 prescription ($p<0.05$). There was no significant effect of prescription for the sounds /ɑ/ and /ʃ/, probably because the spectra of these sounds are dominated by frequencies for which there is little difference between the gains for the different prescription methods.

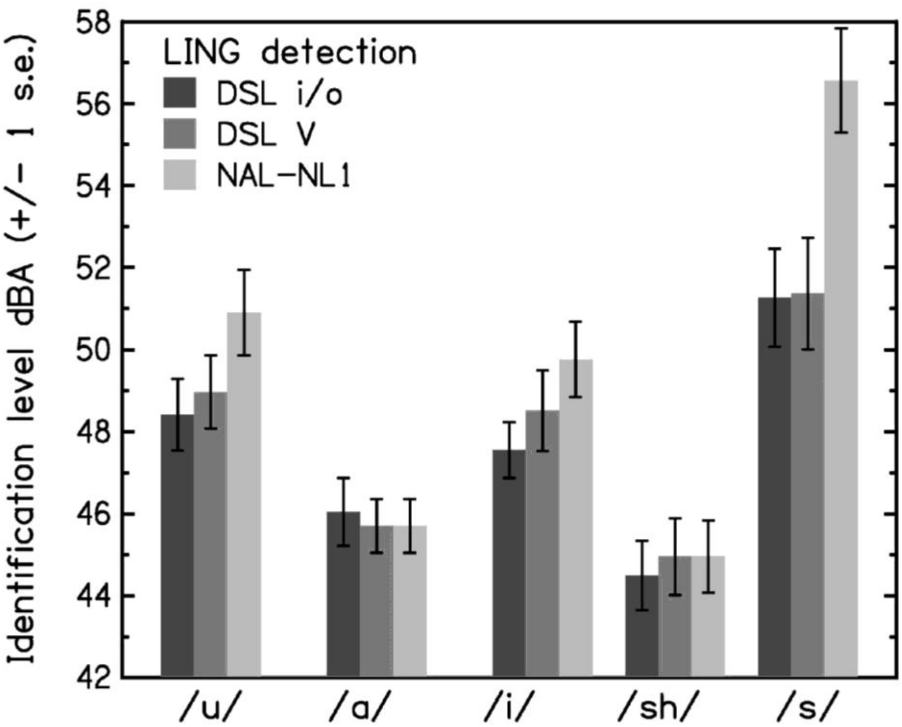

Figure 6. Mean identification level (dBA) for the Ling sounds. Each set of bars shows results for each prescription formula for one of the five sounds. The sounds are specified in terms of their orthographic representations.

Summary

In summary, the results show that:

(1) Closed-set consonant discrimination (CCT) was significantly better with DSL V and DSL i/o than with NAL-NL1. The benefits were predominantly produced by improved perception of place of articulation.

(2) Closed-set discrimination of consonants (CAPT) was significantly better with DSL i/o and DSL V than with NAL-NL1. The improvements occurred for both manner and place of articulation cues.

(3) Closed-set discrimination of vowels in noise (CAPT) was significantly better for DSL V than for the other two prescriptions. The improvements occurred for place and height cues, suggesting that the first two formant frequencies were better perceived using the DSL V prescription formula.

(4) Closed-set detection of consonants (CAPT) was significantly better with DSL i/o and DSL V than with NAL-NL1. This again was predominantly due to better manner and place of articulation perception.

(5) Recognition of open-set words (CAWL) at 50 dBA was significantly better with DSL i/o and DSL V than with NAL-NL1. Scores did not differ significantly across prescriptions for the 65 dBA presentation level. This suggests that when audibility is high, the CAWL test is not sufficiently sensitive to reveal differences across prescriptions.

(6) The SRT estimated using the CPhT test was significantly higher (worse) for NAL-NL1 than for DSL [i/o] or DSL V, indicating that the NAL-NL1 prescription is less effective than the DSL prescriptions in making low-level sounds intelligible.

(7) Identification thresholds for the Ling 5 sounds were significantly higher (worse) for NAL-NL1 than for DSL [i/o] or DSL V. This was particularly the case for the /s/ sound.

DISCUSSION

One goal of this research was to determine if differences in gain recommended by different hearing aid prescription methods would lead to measurable differences in speech recognition performance. Floor and ceiling effects were avoided by the use of age-appropriate speech test materials and, for some of the closed-set tests, by the selection of presentation levels based on the unaided pure-tone thresholds. This allowed meaningful results to be obtained for all age groups. It should be noted that in everyday life, the overall level of speech can vary over a wide range of levels from about 45 dBA to over 85 dBA (Olsen 1998). Differences between prescriptions may not occur over the whole range of levels. For children with severe hearing loss, performance at low levels may be very poor for all prescription methods, while for children with mild hearing loss, performance at medium and high levels may be very good for all prescription methods.

The results showed that, for the whole age range tested (2 to 9 years), the DSL prescription methods led to better detection and discrimination of low-level speech sounds than NAL-NL1, presumably as a consequence of the higher gains and compression ratios recommended by the DSL prescriptions. Additionally, the DSL prescriptions did not lead to lower vowel discrimination scores when the target words were presented in speech-shaped noise, despite the increased low-frequency gains recommended by the DSL prescriptions, which potentially could have increased the “upward spread of masking”. Indeed, DSL V led to better vowel discrimination scores than NAL-NL1 and DSL i/o, as shown in Figure 3. This may indicate that DSL V prescribes gains that lead to a better balance between audibility and upward spread of masking than DSL i/o. The superiority of DSL V over DSL i/o may reflect the fact that DSL V prescribes somewhat lower low-level gains at 500 and 1000 Hz than DSL i/o. The lower gains may help to preserve the relative levels of the first and second formants, which may lead to improved vowel identification.

The higher compression ratios for children with more severe hearing loss might be expected to lead to poorer performance. However, further analysis of the data did not show such an effect. For example, for the CAPT consonant discrimination task, for which the level of the stimuli increased with increasing hearing loss to compensate for effects of audibility, performance was actually somewhat better for the children with severe loss than for the children with moderate loss. It is possible that the children with more severe hearing loss who took part in this study were “high achievers” whose performance was better than average for children with the same amount of hearing loss. However the children were recruited from a wide range of audiology departments and represented all socio-economic groups, so this seems unlikely.

For the open-set tests, the presentation level did not depend on the hearing loss of the individual child. For the open-set CAWL words presented at 65 dBA, which is comparable to the level of conversational speech, there was no significant difference between results for the different prescription methods. However there was a significant effect of severity of hearing loss. This may have happened because, for the level of 65 dBA, performance was mainly limited by supra-threshold factors, such as reduced frequency selectivity (Glasberg & Moore 1986) and reduced sensitivity to temporal fine structure (Hopkins & Moore 2007; Moore 2014), rather than by limited audibility. For CAWL words presented at 50 dBA, which is comparable to the level of the speech of a teacher heard at the back of a classroom or of a parent talking from an adjacent room, the NAL-NL1 prescription led to significantly lower scores than for the DSL prescriptions. This indicates that, to reveal differences between prescriptions, it is important to choose an appropriate presentation level.

At present, hearing aid fittings for children are commonly verified by assessing whether the hearing aid gains match the targets for a specific prescription formula, and functional verification is rarely used. However, additional and perhaps more useful information can be obtained through the use of speech tests of the type described here, and especially by analysis of the transmission of phonetic features. This can provide insight into what cues are being transmitted and help in understanding the effects of hearing-aid signal processing such as multi-channel compression and frequency lowering. It can also help in the evaluation of the benefits of features such as extended bandwidth (Stelmachowicz et al. 2001).

It is possible that the outcomes were somewhat influenced by the fitting that each child was familiar with at the start of the study. All the children were experienced hearing aid users, and had theoretically been fitted using a version of DSL or NAL-NL1 by their local audiology team. However, as noted earlier, the initial assessments of hearing aid output did not reveal fittings that could be clearly identified as corresponding to a specific prescription type, perhaps because insufficient care was taken to adjust the aids to meet targets or because earmoulds had been changed since the initial fittings were made. This is consistent with the findings of several studies that a substantial proportion of children fitted with a specific prescription target had measured aided outputs of their hearing aids whose root-mean-square deviation from the target values was more than 5 dB (McCreery et al. 2013; 2016; Ching et al. 2015). Although the fitting of each child prior to taking part in our study may have influenced the outcomes, we think that that any carry-over effects were probably small, because of the two to four weeks acclimatization/familiarization that was given with a specific fitting before the speech tests were administered.

Another potential issue is that testing took place in several different clinics, and the exact listening conditions varied somewhat across clinics. However, the conditions were consistent for each child, who acted as their own control, and all test sites had sound-treated environments. Also, since the tester was blind to the condition being tested, there was no possibility of biases occurring at the different test sites.

In this paper, we have presented only mean scores and standard errors for each prescription and test, mainly focussing on scores for relatively low sound levels. Of course, other factors must be taken into account when assessing hearing aid prescription procedures. For example, the higher gains for the DSL procedures relative to NAL-NL1 result in greater loudness, and this may lead to loudness tolerance problems with medium and high-level sounds. At the end of our study, each child was allowed to choose which prescription they wanted to be programmed into their own hearing aids. Four of the children with moderate flat hearing loss chose not to have either of the DSL prescriptions, even when their speech discrimination scores were higher than for NAL-NL1, because they found the loudness in some listening situations to be too high with the DSL prescriptions.

Loudness has been considered as a factor in some previous studies of fitting methods conducted with children aged 6 years and above (Crukley & Scollie 2012; Ching et al. 2013). Crukley and Scollie (2012) compared two versions of DSL V, one intended for listening in quiet and one intended for listening in noise. The latter aimed to reduce overall loudness by decreasing gain. As expected, loudness ratings for input levels above 72 dB SPL were significantly lower with the noise prescription. The noise prescription led to a small (4%) but significant reduction in the recognition of consonants in quiet at 50 dB SPL, but no difference at 70 dB SPL or for sentences in noise. These results suggest that it might be beneficial to have a special program for listening in noise, perhaps selected automatically by the hearing aid. Ching et al. (2013) compared groups of children fitted with either NAL or DSL prescriptions. Parents' ratings of loudness discomfort were not significantly different between the two prescription groups. Further research is needed to assess the relative importance of audibility, intelligibility, loudness, and sound quality in determining overall preference for and benefit from different prescription methods.

Although further research is needed, one clinical implication of the present results is that speech testing for evaluating the effectiveness of hearing aids may be conducted with children as young as two years old. Furthermore, the speech tests can be pre-recorded and/or run via computer, avoiding the variability and biases associated with live-voice testing. Given adequate test-retest reliability, the outcomes of the tests could potentially be used for identifying problems and for fine tuning of hearing aid fittings based on phoneme confusions. For example, frequent confusions of fricatives might indicate a need for more high-frequency gain or less low-frequency gain. The speech tests could potentially allow monitoring of the development of auditory and speech-perceptual skills, with the goal of indicating when further intervention might be needed. Speech testing for young children is not an alternative to obtaining measures of aided gain and the Aided SII; audibility is critical but not sufficient to ensure adequate speech perception. Rather, speech testing provides additional important information that can be helpful in adjusting hearing aids to optimize benefit.

CONCLUSIONS

Using age-appropriate closed-set and open-set speech tests, designed to avoid floor and ceiling effects, we found significant differences between scores for the different hearing aid prescription methods. The higher output levels prescribed by the DSL i/o and DSL V prescription methods relative to NAL-NL1 led to significantly better detection and discrimination of low-level sounds. However, open-set speech recognition testing at 65 dB did not reveal differences between prescription methods, probably because performance was mainly limited by supra-threshold factors rather than by audibility.

ACKNOWLEDGEMENTS

We thank Deafness Research UK (now part of Action on Hearing Loss) for funding, Phonak and Oticon for the study hearing aids, PC Werth and Audioscan for the Verifit system, John Deeks for help with calibration, the children and their families, and the audiology host centers. We also thank Ben Hornsby and three reviewers for helpful comments on an earlier version of this paper.

REFERENCES

ANSI (1996). *ANSI S3.22-1996, Specification of hearing aid characteristics*. New York: American National Standards Institute.

ANSI (1997). *ANSI S3.5-1997. Methods for the calculation of the speech intelligibility index*. New York: American National Standards Institute.

Bagatto, M., Scollie, S. D., Hyde, M., et al. (2010). Protocol for the provision of amplification within the Ontario infant hearing program. *Int J Audiol*, *49 Suppl 1*, S70-79.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*, 307-310.

Byrne, D., Dillon, H., Ching, T., et al. (2001). NAL-NL1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures. *J Am Acad Audiol*, *12*, 37-51.

Ching, T. Y., Dillon, H., Hou, S., et al. (2013). A randomized controlled comparison of NAL and DSL prescriptions for young children: hearing-aid characteristics and performance outcomes at three years of age. *Int J Audiol*, *52 Suppl 2*, S17-28.

Ching, T. Y., Dillon, H., Katsch, R., et al. (2001). Maximizing effective audibility in hearing aid fitting. *Ear Hear*, *22*, 212-224.

Ching, T. Y., & Hill, M. (2007). The Parents' Evaluation of Aural/Oral Performance of Children (PEACH) scale: normative data. *J Am Acad Audiol*, *18*, 220-235.

Ching, T. Y., Hill, M., & Dillon, H. (2008). Effect of variations in hearing-aid frequency response on real-life functional performance of children with severe or profound hearing loss. *Int J Audiol*, *47*, 461-475.

Ching, T. Y., Quar, T. K., Johnson, E. E., et al. (2015). Comparing NAL-NL1 and DSL v5 in hearing aids fit to children with severe or profound hearing loss: Goodness of fit-to-targets, impacts on predicted loudness and speech intelligibility. *J Am Acad Audiol*, *26*, 260-274.

Ching, T. Y., Scollie, S. D., Dillon, H., et al. (2010a). A cross-over, double-blind comparison of the NAL-NL1 and the DSL v4.1 prescriptions for children with mild to moderately severe hearing loss. *Int J Audiol*, *49 Suppl 1*, S4-15.

Ching, T. Y., Scollie, S. D., Dillon, H., et al. (2010b). Evaluation of the NAL-NL1 and the DSL v.4.1 prescriptions for children: Paired-comparison intelligibility judgments and functional performance ratings. *Int J Audiol*, *49 Suppl 1*, S35-48.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd Ed.)*. Hillsdale, NJ: Erlbaum.

Cornelisse, L. E., Seewald, R. C., & Jamieson, D. G. (1995). The input/output formula: A theoretical approach to the fitting of personal amplification devices. *J Acoust Soc Am*, *97*, 1854-1864.

Crukley, J., & Scollie, S. D. (2012). Children's speech recognition and loudness perception with the Desired Sensation Level v5 Quiet and Noise prescriptions. *Am J Audiol*, *21*, 149-162.

Davidson, L. S., & Skinner, M. W. (2006). Audibility and speech perception of children using wide dynamic range compression hearing aids. *Am J Audiol*, *15*, 141-153.

Dillon, H. (1996). Compression? Yes, but for low or high frequencies, for low or high intensities, and with what response times? *Ear Hear*, *17*, 287-307.

Glasberg, B. R., & Moore, B. C. J. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *J Acoust Soc Am*, *79*, 1020-1033.

Glista, D., Scollie, S., Moodie, S., et al. (2014). The Ling 6(HL) test: typical pediatric performance data and clinical use evaluation. *J Am Acad Audiol*, *25*, 1008-1021.

Govaerts, P. J., Daemers, K., Yperman, M., et al. (2006). Auditory speech sounds evaluation (A§E®): a new test to assess detection, discrimination and identification in hearing impairment. *Cochlear Implants Int*, *7*, 92-106.

Hacker, M. J., & Ratcliff, R. (1979). A revised table of d' for M-alternative forced choice. *Percept Psychophys*, *26*, 168-170.

Hopkins, K., & Moore, B. C. J. (2007). Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *J Acoust Soc Am*, *122*, 1055-1068.

Keidser, G., Dillon, H., Flax, M., et al. (2011). The NAL-NL2 prescription procedure. *Audiol Res*, *1:e24*, 88-90.

Lovett, R., Summerfield, Q., & Vickers, D. (2013). Test-retest reliability of the Toy Discrimination Test with a masker of noise or babble in children with hearing impairment. *Int J Audiol, 52*, 377-384.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide, 2nd Ed.* New York: Erlbaum.

Marriage, J. E., & Moore, B. C. J. (2003). New speech tests reveal benefit of wide-dynamic-range, fast-acting compression for consonant discrimination in children with moderate to severe hearing loss. *Int J Audiol*, *42*, 418-425.

Marriage, J. E., Moore, B. C. J., Stone, M. A., et al. (2005). Effects of three amplification strategies on speech perception by children with severe and profound hearing loss. *Ear Hear*, *26*, 35-47.

McConkey Robbins, A., Koch, D. B., Osberger, M. J., et al. (2004). Effect of age at cochlear implantation on auditory skill development in infants and toddlers. *Arch Otolaryngol Head Neck Surg*, *130*, 570-574.

McCreery, R., Walker, E., Spratford, M., et al. (2016). Stability of audiometric thresholds for children with hearing aids applying the American Academy of Audiology Pediatric Amplification Guideline: Implications for safety. *J Am Acad Audiol*, *27*, 252-263.

McCreery, R. W., Bentler, R. A., & Roush, P. A. (2013). Characteristics of hearing aid fittings in infants and young children. *Ear Hear*, *34*, 701-710.

McCreery, R. W., & Stelmachowicz, P. G. (2011). Audibility-based predictions of speech recognition for children and adults with normal hearing. *J Acoust Soc Am*, *130*, 4070-4081.

Meinzen-Derr, J., Wiley, S., Creighton, J., et al. (2007). Auditory Skills Checklist: clinical tool for monitoring functional auditory skill development in young children with cochlear implants. *Ann Otol Rhinol Laryngol*, *116*, 812-818.

Moore, B. C. J. (2007). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues, 2nd Ed.* Chichester: Wiley.

Moore, B. C. J. (2008). The choice of compression speed in hearing aids: Theoretical and practical considerations, and the role of individual differences. *Trends Amplif*, *12*, 103-112.

Moore, B. C. J. (2014). *Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss*. Singapore: World Scientific.

Moore, B. C. J., Füllgrabe, C., & Stone, M. A. (2011). Determination of preferred parameters for multi-channel compression using individually fitted simulated hearing aids and paired comparisons. *Ear Hear, 32*, 556-568.

Moore, B. C. J., & Sek, A. (2013). Comparison of the CAM2 and NAL-NL2 hearing-aid fitting methods. *Ear Hear, 34*, 83-95.

Mueller, H. G., Hawkins, D. B., & Northern, J. L. (1992). *Probe Microphone Measurements: Hearing Aid Selection and Assessment*. San Diego, CA: Singular.

Olsen, W. O. (1998). Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report. *Am J Audiol*, *7*, 21-25.

Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *J Acoust Soc Am*, *83*, 2322-2327.

Quar, T. K., Ching, T. Y., Newall, P., et al. (2013). Evaluation of real-world preferences and performance of hearing aids fitted according to the NAL-NL1 and DSL v5 procedures in children with moderately severe to profound hearing loss. *Int J Audiol*, *52*, 322-332.

Renfrew, C. (1995). *Word Finding Vocabulary Test*. Oxford: Winslow Press.

Robbins, A., Renshaw, J., & Osberger, M. J. (1988). *Common Phrases Test*. Indianapolis: Indiana University School of Medicine.

Scollie, S. D. (2008). Children's speech recognition scores: the Speech Intelligibility Index and proficiency factors for age and hearing level. *Ear Hear*, *29*, 543-556.

Scollie, S. D., Glista, D., Tenhaaf, J., et al. (2012). Stimuli and normative data for detection of Ling-6 sounds in hearing level. *J Am Acad Audiol*, *21*, 232-241.

Scollie, S. D., Seewald, R. C., Cornelisse, L., et al. (2005). The Desired Sensation Level multistage input/output algorithm. *Trends Amplif*, *9*, 159-197.

Stelmachowicz, P. G., Hoover, B. M., Lewis, D. E., et al. (2000). The relation between stimulus context, speech audibility, and perception for normal-hearing and hearing-impaired children. *J Speech Lang Hear Res*, *43*, 902-914.

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., et al. (2001). Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults. *J Acoust Soc Am*, *110*, 2183-2190.

Stiles, D. J., Bentler, R. A., & McGregor, K. K. (2012a). The Speech Intelligibility Index and the pure-tone average as predictors of lexical ability in children fit with hearing aids. *J Speech Lang Hear Res*, *55*, 764-778.

Stiles, D. J., McGregor, K. K., & Bentler, R. A. (2012b). Wordlikeness and word learning in children with hearing loss. *Int J Lang Commun Disord*, *48*, 200-206.

Stone, M. A., & Moore, B. C. J. (1992). Syllabic compression: Effective compression ratios for signals modulated at different rates. *Br J Audiol*, *26*, 351-361.

Stone, M. A., & Moore, B. C. J. (2004). Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task. *J Acoust Soc Am*, *116*, 2311-2323.

Strauss, S., & van Dijk, C. (2008). Hearing instrument fittings of pre-school children: do we meet the prescription goals? *Int J Audiol*, *47 Suppl 1*, S62-71.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res*, *28*, 455-462.

Thornton, A. R., & Raffin, M. J. (1978). Speech-discrimination scores modeled as a binomial variable. *J Speech Hear Res*, *21*, 507-518.

Vickers, D. A., Moore, B. C. J., Majeed, A. A., et al. (2018). Closed set speech discrimination tests for assessing young children. *Ear Hear*, (submitted).