Using Speech Perception Measures to Guide the Choice of Amplification

Josephine Marriage, Deborah Vickers, Thomas Baer and Brian C.J. Moore

Introduction

The early fitting of infants with hearing aids relies largely on the verification of amplification characteristics using simulated real ear measurements (REM) that are matched to prescription targets derived from audiometric thresholds. The audibility of different parts of the speech spectrum can be inferred from the audiometric thresholds and the estimated applied gain, and is expressed as the speech intelligibility index (SII; ANSI 1997). However, the fidelity of the representation of important temporal and spectral cues within speech is generally not assessed, and detection and discrimination of speech features for the infant or child remain unconfirmed. Thus REM allows verification that the hearing aid gains fit the target, but not the verification of functional benefit for speech intelligibility. The latter requires speech testing, with the form of the test(s) chosen so as to be age-appropriate and to avoid floor and ceiling effects.

Rationale for the Present Study

The literature on early auditory development for normally hearing infants indicates a sequence in which different classes of speech cues acquire meaning (or become useful) at different times (Werner and Marean 1996). In the early months of life, for example, the child may attend to voicing patterns that characterize and identify the talker, rather than using those patterns to aid the recognition of words or phrases. For normal-hearing infants there is preferential perception and neural encoding of phonemic contrasts from the child's own language by the second half of the first year of life (Kuhl et al. 2008).

Taking this developmental sequence into account, it seems likely that the relative importance of different frequency bands may change depending on whether a child is in the pre- or peri-lingual period of communication. For example, when a child is first becoming aware of human voices, access to low-frequency voicing cues (in the range 200–600 Hz) will be more important than access to the high-frequency cues (between 2000 and 3000 Hz) that are important for some consonant contrasts. Thus, the frequency-importance function used to calculate the SII, which specifies the relative importance of different frequency bands for adults with normal hearing, may not be applicable to infants and children, especially when they have hearing loss. However, a practical problem in verifying the benefits of amplification of low frequencies is that techniques for estimating hearing thresholds for infants and young children (e.g., auditory brainstem responses [ABR] and auditory steady-state responses [ASSR]) are less precise for low frequencies than for medium and high frequencies (Stapells 1994; Rance et al. 2005; Van Maanen and Stapells 2009).

For the hearing-impaired child, the appropriateness of hearing aid amplification characteristics will determine the degree to which environmental and speech sounds are audible, and this will in turn affect whether auditory development can proceed in a normal or near-normal manner. In order to capitalize on early neural plasticity, close and continuous monitoring of auditory awareness is required. The hearing aid prescription can be set initially and then fine-tuned using three sources of information. The initial fitting relies largely on estimates of hearing thresholds obtained using ABR (supple-

---

Address correspondence to: Josephine Marriage, Ph.D., Research Associate, University of Cambridge, Director, Chear, 30 Fowlmere Road, Shepreth, Royston, Herts SG8 6QS UK, Email: Jem1003@cam.ac.uk.

mented or not by ASSR data – see the Chapter by Stapells). Over the next few months accurate behavioral hearing responses can usually be derived, and these should supersede threshold estimates based on ABR/ASSR. But both of these should be supplemented by close observation and monitoring of the child's auditory awareness and developing communication responses. This functional information allows evaluation of the child's progress between clinical assessments and specifies targets for listening and vocalization support. In the present study, we describe the outcomes of a variety of tests that can be used at different stages of development to assess functional performance in terms of awareness of speech and speech discrimination and identification. The tests are used to compare the effectiveness of three fitting procedures in restoring the audibility of soft speech sounds. Although children with a range of ages were tested, we present here only the data for all children tested using a given type of test. Effects of age will be presented in a subsequent publication. The fitting procedures used in this study are discussed next.

Hearing Aid Prescriptions for Infants and Children

Any research assessing the effectiveness of hearing aid prescription methods with young children must minimize the possibility that any of the prescription conditions will have a negative impact on the child, even for a limited period of time. Therefore only published and professionally accepted amplification procedures can be used. There are three prescription methods that are widely used for fitting hearing aids to children around the world: the two versions of DSL (DSL[i/o] and the updated version 5; Cornelisse, Seewald and Jamieson 1995; Scollie et al. 2005) and NAL-NL1 (Byrne, Dillon, Ching, Katsch and Keidser 2001). Each child needs a period of at least one week and preferably two weeks use of hearing aids fitted using a given prescription method prior to being tested; this allows the child to become familiar with the nature of sounds provided by the prescription. There are differences between the rationales and amplification characteristics for NAL-NL1 and the DSL prescriptions: NAL-NL1 generally prescribes less low- and high-frequency gain than the DSL methods, particularly for severe or profound hearing loss, as shown in figure 1. NAL-NL1 also prescribes less compression than the DSL methods, especially at high frequencies. The amount of compression in hearing aids represents a compromise. Compression is required to ensure that low-level sounds are audible while intense sounds are not uncomfortably loud (Villchur 1973; Moore 2008). The amount of compression required to achieve this generally increases with increasing hearing loss and the consequent reduced dynamic range. However, if too much compression is applied, this introduces undesirable effects such as reduced modulation depth in speech and reduction of spectral contrast. 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 there is a compromise between increased audibility of speech cues and increased susceptibility to the effects of noise. Compression at high frequencies has been found to be beneficial for some hearing-impaired adults (Laurence, Moore and Glasberg 1983; Moore, Johnson, Clark and Pluvinage 1992; Marriage and Moore 2003), but the results for children are less clear-cut (Marriage, Moore, Stone and Baer 2005). A collaborative study comparing the DSL and NAL-NL1 prescriptions for older children showed no clear overall benefit for one prescription over the other (Ching, Scollie, Dillon and Seewald 2010a; Ching et al. 2010b).

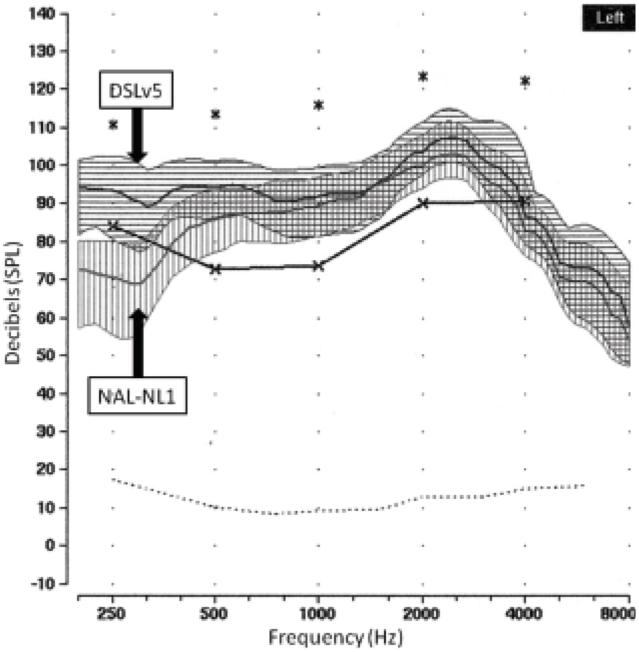

Figure 1. Sample amplification characteristics for NAL-NL1 and DSLv5 for a moderate to severe hearing loss.

Speech Test Development and Selection

An assessment of functional hearing for speech requires test materials that can be used to derive reliable results for hearing-impaired children as young as 2 years of age. The tests used in this study included the minimum detection level for repetition of individual phonemes, word discrimination, word recognition, phrase or sentence identification, and speech in noise understanding. Although some available speech tests were suitable for our purposes, new materials or methods of presentation were developed and adapted for use with hearing-impaired children from pre-school age to infant school age. The final test battery included the following: (1) three closed-set monosyllabic speech tests with pictures, designed for use at different stages of linguistic development. Pictures were presented on a touch sensitive screen or laptop with mouse; (2) ten open-set word lists for use in quiet and/or speech-shaped noise; (3) Ling five stimuli (phonemes /u, a, i, sh, s/; Glista et al. 2009); and (4) the pre-recorded common phrase test (CPT; Osberger et al. 1991).

These tests furnish information on the functional effect of hearing-aid fitting for: detection of speech sounds with spectra dominated by low or high frequencies; discrimination between vowels; discrimination between consonants; recognition of words; and understanding of running speech including effects of masking by noise. In addition, informal observations of listening effort by the child and the opinion of the child on the quality and comfort of sound, including perception of noise, were obtained; however, these data are not reported here. All classes of speech sounds were considered to be important in this study and were represented in the speech tests.

Norms for the speech test battery were derived from 36 normally hearing (NH) children between the ages of 2 and 8 years. Levels were calibrated using a speech-shaped noise that was presented via an audiometer and loudspeaker. The level was measured using a sound level meter at the position of the child's head. The minimum level at which the closed-set speech tests could be performed reliably depended somewhat on the specific testing room used, but was typically about 30 dBA for all age groups of the NH children. The Renfrew word finding test (Renfrew 1995) was used as a vocabulary screen to determine which individual parts of the speech test battery to use with each child and to maintain an appropriate level of challenge and thereby self-motivation. The test-retest repeatability of the closed-set tests was defined as the average difference (regardless of sign) between scores obtained on two different test sessions. The repeatability for children with normal hearing in the age group 4-8 years was found to be 7.6%.

Examples of the Closed-set Speech Material

In the closed-set tests (1) to (3) listed below, the target word was presented from a loudspeaker positioned directly in front of the child at a distance between 0.6 and 1 meter. Four pictures were presented on a touch-sensitive screen, one of which corresponded to the target item. The child responded by touching one of the pictures, or clicking on it with a mouse. The only factor that was varied was the hearing aid prescription in the child's hearing aids (NAL-NL1, DSL[i/o], DSLv5). Although a number of test venues were used over the course of the study, each child was tested in the same room across all testing sessions. The test measures included:

(1) Identification of phonemes in a closed-set four-alternative task with pictures (2-8 years). Examples: *eye, ice, lice, slice, or why, wine, eye, wise.*

(2) Consonant discrimination with closed-set testing with a four-item picture task in which the vowel was the same. For younger ages, both the initial and final consonants within each group of four items were different, providing more contrastive speech cues. Examples: *three, key, sheep, feet or hen, peg, egg, bed.* For older ages, the items differed either in the initial consonant or the final consonant, but not both. Examples: *fat, cat, bat, mat* (word-initial) or *cheese, cheat, cheap, cheek* (word-final).

(3) Vowel discrimination with 12 or 20 item versions, presented in noise or quiet. Example: *tea, tie, tar, two.*

(4) Ling five-repetition task for pre-recorded phonemes: /u, a, i, sh, s/. For this task, the child was asked to repeat the speech sound that was heard. The level of all sounds was varied, and the sounds that were correctly identified at each level were recorded. The score reported is the lowest level at which a given sound could be reliably identified.

Speech Presentation Levels

In order to assess performance while avoiding floor or ceiling effects, the presentation level in dBA of the closed-set test material was individually set using the following formula:

Level = [(3-frequency audiometric thresholds (PTA) in better hearing ear) 0.4] + 30 dB For example, if the PTA was 40 dB, the level was set to 16 + 30 = 46 dB SPL.

The three frequencies were chosen individually for each child to be the frequencies for which the audiometric thresholds were highest.

This formula reflects the fact that, as hearing loss increases, more compression is required to restore the audibility of weak sounds while preventing excessive loudness and/or distortion in the hearing aid from intense sounds. Since too much compression can have deleterious effects, as discussed above, the lowest sound level for which audibility can be restored needs to increase with increasing hearing loss.

For the open-set materials presented in quiet, testing was conducted at 50 and 65 dBA. For the open-set materials presented in speech-shaped noise, the speech level was set to 60 dBA and the noise level was set 5 or 10 dB below this.

Subjects and Test Conditions

Fifty-four children were initially enrolled into the study, and 44 children with moderate and severe bilateral hearing loss completed all conditions. The children were divided into three age groups:

Group 1 (2–3 yrs): n = 8
Group 2 (4–5 yrs): n = 14
Group 3 (6–9 yrs): n = 22

Of the ten children who dropped out of the study, only one dropped out due to an inability to complete the speech testing. There was some indication of wider communication difficulties for this child.

Several hearing aid types were used for the study. Hearing aids were chosen to be compatible with radio aid equipment that the child was using at the time of enrolment. The hearing aids used were Phonak Savia Art®, Oticon Safran®, and Phonak Naida®. Two children were tested using their own Oticon Spirit II® hearing aids, as they did not want to change their hearing aid model.

Hearing aid gains were adjusted to match targets for NAL-NL1, DSL[i/o] and DSLv5, the gains for each method being stored in a different program in the hearing aid, using a code that was unknown to the tester – that is, the tester was blind as to which program number corresponded to a given prescription. Gains were verified with real-ear-to-coupler difference (RECD) measures using real speech input with the Verifit® REM system. The different prescriptions were used in a balanced randomized order across subjects. Subjects wore the study hearing aids with each prescription in turn, typically for between two and four weeks for each prescription. At the end of this acclimatization period, they were assessed using the speech test battery and the next prescription was activated. The tester was blind to the prescription fitting condition at the time of testing.

Results

Ling Sound Detection Level in dBA

Group results for all subjects for the lowest level at which the child was able to repeat the Ling five sounds correctly are shown in table 1. The numbers given are the mean level in dBA required for correct repetition.

A separate within-subject one-way analysis of variance (ANOVA) with factor prescription type was conducted across all subjects for each Ling sound. The p value associated with each sound is given in the rightmost column of Table 1. No differences between prescriptions were found for /a/ and /sh/ detection levels.

Ling sounds	DSLv5	DSL [i/o]	NAL-NL1	F value (df)	*p* value
/u/	38.2	38.9	40.8	4.20 (2, 88)	$p = 0.018$
/i/	36.5	37.2	38.8	4.13 (2, 88)	$p = 0.019$
/s/	43.5	43.7	48.6	17.3 (2, 88)	$p < 0.001$
/a/	40.6	40.5	40.7	0.12 (2, 88)	$p = 0.89$
/sh/	40.5	41.2	41.3	0.84 (2, 88)	$p = 0.43$

Table 1. Mean scores for the Ling sounds across all subjects. The final two columns show F and p values for one-way within-subject ANOVAs with factor prescription type.

Post hoc tests, based on Fisher's protected least significant differences test, showed that the levels required for correct repetition of /u/, /i/ and /s/ were higher for NAL-NL1 than for DSLv5 and DSL [i/o]. The levels did not differ significantly for DSLv5 and DSL [i/o]. In summary, the DSL prescriptions allowed the Ling sounds to be correctly identified at lower levels than the NAL-NL1 prescription.

Closed-set 2-8 years	Score type	DSL v5	DSL [i/o]	NAL-NL1	F value (df)	*p* value
Consonant discrimination	% correct	80.1	81.8	74.1	15.1 (2,88)	*p* <0.001
Phoneme identification in word	% correct	84.2	95.6	77.9	10.9 (2,60)	*p*<0.001
Vowel disc in noise	% correct 0 dB SNR	84.0	86.5	81.2	1.14 (2,60)	*p* = 0.32
Open set 4–8 years						
Word recog. (50 dBA)	Score /30	22.4	23.1	19.7	9.68 (2,66)	*p* <0.001
Word recog. (65 dBA)	Score/30	26.7	26.5	26.3	0.33 (2,70)	*p* = 0.72
Phrase test (CPT)	dBA (adaptive presentation)	39.4	38.8	41.5	7.48 (2,70)	*p*=0.001
Word recognition in noise	Score /30 5 or 10 dB SNR	24.5	24.5	21.8	3.32 (2,22)	*p* = 0.055

Table 2. Results of the closed- and open-set speech tests. The first column identifies the test. The second column indicates the type of score. The third to fifth columns indicate the prescription method used. The final two columns show F and p values for within-subject ANOVAs, based on the data for all subjects.

Closed and Open Set Speech Tests

Results for the closed and open set speech tests are shown in Table 2. The type of score is indicated in the second column. Open-set speech testing was usually not feasible with the younger group of subjects for the open-set speech materials used here.

The results show:

(1) Consonant discrimination was about seven percentage points better with DSLv5 and DSL [i/o] than with NAL-NL1.
(2) Identification of phonemes in words was highest for DSL [i/o] and lowest for NAL-NL1.
(3) Discrimination of vowels in noise did not differ significantly across prescriptions.
(4) Recognition of words at 50 dBA was highest for DSL [i/o] and lowest for NAL-NL1.
(5) Recognition of words at 65 dBA showed no significant difference between prescriptions.
(6) The "threshold" level for the CPT test was higher for NAL-NL1 than for DSLv5 or DSL [i/o].
(7) Recognition of words in noise was slightly lower for NAL-NL1 than for DSLv5 and DSL [i/o], but the difference just failed to reach statistical significance.

Discussion

These results show better speech discrimination performance for the DSL prescriptions than for the NAL-NL1 prescription, for both consonants and vowels presented in quiet. Presumably this is a consequence of the lower gains and compression ratios recommended by NAL-NL1, which would have led to reduced audibility of low-level sounds. Additionally, the DSL prescriptions did not lead to lower discrimination or recognition scores when target words were presented in speech-shaped noise, despite the increased low-frequency gains recommended by the DSL prescriptions. Such lower discrimination for the DSL prescriptions might have been expected if the greater low-frequency gains for those prescriptions led to greater upward spread of masking or greater distraction. Rather, scores in noise were slightly, but just non-significantly, higher for the DSL prescriptions than for NAL-NL1.

Although there was a slight trend for performance to improve with test familiarity (i.e., with test order), the effects were not significant for any test. This suggests that learning effects with these tests are small, at least over the short term.

In this brief report, we have presented only mean scores for each prescription and test, mainly focusing on scores for relatively low sound levels. Of course, other factors must be taken into account when assessing prescriptive procedures. For example, the higher gains for the DSL procedures relative to NAL-NL1 result in greater loudness, and this may lead to tolerance problems with medium and high-level sounds. Loudness preferences for each prescription will be included in forthcoming publications, which will also include analyses of patterns of confusion, age effects, and individual differences.

Conclusions

The DSL methods prescribe more gain and more compression than NAL-NL1. Using age-appropriate closed-set and open-set speech tests, designed to avoid floor and ceiling effects, we have shown that the higher gains for the DSL methods lead to significantly improved detection and discrimination of low-level sounds. However, open-set speech recognition testing at 65 dB did not reveal differences between prescription methods, presumably because performance was limited by supra-threshold factors (such as frequency selectivity, temporal resolution, and sensitivity to temporal fine structure) rather than by audibility. In future publications we will consider whether the benefits of the DSL procedures demonstrated in this study are offset by factors such as higher loudness (than for NAL-NL1) for high input levels.

Acknowledgements

Thanks to Deafness Research UK for funding (COREC No 06/Q0108/321), to Phonak and Oticon for the study hearing aids, to PCWerth and Audioscan for the Verifit®, and to the children and families and the NHS audiology host centres.

References

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

Byrne, D., Dillon, H., Ching, T., Katsch, R., and Keidser, G. 2001. NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures. *Journal of the American Academy of Audiology* 12: 37–51.

Ching, T.Y., Scollie, S.D., Dillon, H., and Seewald, R. 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. *International Journal of Audiology* 49 Suppl. 1: S4–15.

Ching, T.Y., Scollie, S.D., Dillon, H., Seewald, R., Britton, L., Steinberg, J., Gilliver, M., and King, K.A. 2010b. Evaluation of the NAL-NL1 and the DSL v.4.1 prescriptions for children: Paired-comparison intelligibility judgments and functional performance ratings. *International Journal of Audiology* 49 Suppl. 1: S35–48.

Cornelisse, L.E., Seewald, R.C., and Jamieson, D.G. 1995. The input/output formula: A theoretical approach to the fitting of personal amplification devices. *Journal of the Acoustical Society of America* 97: 1854–1864.

Glista, D., Scollie, S., Bagatto, M., Seewald, R., Parsa, V., and Johnson, A. 2009. Evaluation of nonlinear frequency compression: Clinical outcomes. *International Journal of Audiology* 48: 632–644.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. 2008. Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London B* 363: 979–1000.

Laurence, R.F., Moore, B.C.J., and Glasberg, B.R. 1983. A comparison of behind-the-ear high-fidelity linear aids and two-channel compression hearing aids in the laboratory and in everyday life. *British Journal of Audiology* 17: 31–48.

Marriage, J.E., and 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. *International Journal of Audiology* 42: 418–425.

Marriage, J.E., Moore, B.C.J., Stone, M.A., and Baer, T. 2005. Effects of three amplification strategies on speech perception by children with severe and profound hearing loss. *Ear and Hearing* 26: 35–47.

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

Moore, B.C.J., Johnson, J.S., Clark, T.M., and Pluvinage, V. 1992. Evaluation of a dual-channel full dynamic range compression system for people with sensorineural hearing loss. *Ear and Hearing* 13: 349–370.

Osberger, M.J., Miyamoto, R.T., Zimmerman-Phillips, S., Kemink, J.L., Stroer, B.S., Firszt, J.B., and Novak, M.A. 1991. Independent evaluation of the speech perception abilities of children with the Nucleus 22-channel cochlear implant system. *Ear and Hearing* 12: 66S–80S.

Rance, G., Roper, R., Symons, L., Moody, L.J., Poulis, C., Dourlay, M., and Kelly, T. 2005. Hearing threshold estimation in infants using auditory steady-state responses. *Journal of the American Academy of Audiology* 16: 291–300.

Renfrew, C. 1995. *Word finding vocabulary test.* Oxford: Winslow Press.

Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., Beaulac, S., and Pumford, J. 2005. The Desired Sensation Level multistage input/output algorithm. *Trends in Amplification* 9: 159–197.

Stapells, D.R. 1994. Low-frequency hearing and the auditory brainstem response. *American Journal of Audiology* 3: 11–13.

Van Maanen, A., and Stapells, D.R. 2009. Normal multiple auditory steady-state response thresholds to air-conducted stimuli in infants. *Journal of the American Academy of Audiology* 20: 196–207.

Villchur, E. 1973. Signal processing to improve speech intelligibility in perceptive deafness. *Journal of the Acoustical Society of America* 53: 1646–1657.

Werner, L.A., and Marean, G.C. 1996. *Human auditory development.* Boulder, CO: Westview Press.