Hearing tests are just child's play: The Sound Scouts game for children entering school

Harvey Dillon[1, 2], Carolyn Mee[3], Jesus Cuauhtemoc Moreno[3] and John Seymour[1]

1. National Acoustic Laboratories, Sydney, Australia
2. University of Manchester, Manchester, UK
3. Sound Scouts Pty Ltd, Sydney, Australia

Key words: Hearing, screening, children, school, tablet, game, computer-based

Acronyms: 4FAHL (four-frequency average hearing loss), APD (auditory processing disorder), SNR (signal-to-noise ratio), SRTn (speech reception threshold in noise), SRTq (speech reception threshold in quiet), ΔSRTq (difference between two speech reception thresholds in quiet), TRT (tone reception threshold)

Corresponding author: Harvey Dillon

Harvey.dillon@nal.gov.au

National Acoustic Laboratories

16 University Avenue

Macquarie University, 2109

Sydney,

AUSTRALIA

Abstract

Objective

To create a hearing test useable without the involvement of a clinician or calibrated equipment, for children aged 5 or older.

Design

The tablet-based app (Sound Scouts) includes tests of speech in quiet, speech in noise, and tones in noise, all embedded in a game designed to maintain attention. Data were collected to intelligibility-equalize the stimuli, establish normative performance, and evaluate the sensitivity with which Sound Scouts detected known hearing problems and identified their type.

Study sample

Participants were children from age 5 to 14 (394 with normal hearing, 97 with previously identified hearing loss) and 50 adults with normal hearing.

Results

With pass-fail criteria set such that 98% of children with normal hearing passed Sound Scouts, 85% of children with hearing loss failed Sound Scouts (after exclusion of inconclusive or incomplete results). No child with four-frequency average hearing thresholds of 30 dB HL or greater in their poorer ear passed Sound Scouts. Hearing loss type was correctly identified in only two-thirds of those cases where the algorithm attempted to identify a single type of loss.

Conclusions

Sound Scouts has specificity and sensitivity sufficiently high to provide hearing screening from around the time children typically enter school.

Dillon: Sound Scouts hearing test game

Introduction

Universal new-born hearing screening has been successful in detecting children born with bilateral hearing loss of moderate degree or greater, as evidenced by the large number of children fitted with hearing aids in their first year of life (Australian Hearing, 2015). Despite this, in Australia where comprehensive statistics are available covering all children in the country with hearing aids, more children receive their first hearing aids during the first three years of school than receive them during their first year of life (Australian Hearing, 2015). These later-fitted children typically are brought in for a hearing assessment because they are performing badly at school (Nilsson et al, in preparation). All too often, their self-esteem and attitude to school have been adversely affected by the time a parent, teacher or other professional works out that hearing loss might be the cause of their problem and seeks an assessment (Bess et al., 1998; Kuppler et al., 2013). Although they commonly have mild loss in one ear and mild loss or normal hearing in the other ear, of the 431 children in Australia aged 5 or 6 years at the time they received their first hearing aid(s) in 2014, the loss exceeded 40 dB HL three-frequency average loss (3FAHL; average of 500, 1000 and 2000 Hz) for 42% of children in their poorer ear and for 11% of children in their better ear (Australian Hearing, 2015).

This paper describes a computer-based, low-cost solution to hearing screening, which we have called Sound Scouts. It uses mobile technology, with stimuli presented in the format of a game, to engage and retain the child's attention during the test. The idea of building a hearing test into a tablet device or computer together with graphics intended to be attractive to children is not new. However, other implementations that we are aware of for testing hearing thresholds (Yeung et al; 2013) or auditory processing ability (Barry, Ferguson and Moore, 2010) require that a trained clinician carry out the testing using calibrated headphones. Comprehensive hearing screening around school entry age (often around 5 years of age) is more likely to be widely adopted if a solution can be found that does not require testing by clinicians or specialised equipment. Our aim was to devise a solution that could be used at home, in schools, community health centres, or general medical practices, for children as young as 5, or preferably 4 years of age, supervised by adults with no training in hearing testing.

As the ultimate goal is to ensure that children are able to understand their teacher in a typical classroom, we designed the game so that, in addition to detecting hearing loss, it should be sensitive to auditory processing disorders that could impede understanding in noisy classrooms. Detection of sensorineural loss, because of its unquestioned permanence and effect on both audibility and clarity of speech, was however given the highest priority in test design.   We also sought to build in the possibility of determining the most likely cause of hearing difficulties, which necessitated including more than one type of hearing test within the game. The evaluation described in this paper was restricted to children with normal hearing, sensorineural hearing loss, and conductive hearing loss. This paper describes the tests employed in the game, their principles, technical development, normative data, and their ability to detect hearing loss and identify the type of loss.

Study 1: Test description and development

Sound Scouts is an interactive game for tablet devices built around a speech-in-noise task and a tone-in-noise task, preceded by a separate speech-in-quiet calibration game that both provides one of the test metrics and is also used to set the level of the stimuli for the main game. Each of these is described in the following sections, and a video demonstration of them is available at https://www.youtube.com/watch?v=ybpkuiJ6i6Y.

Dillon: Sound Scouts hearing test game

Calibration game - speech in quiet

Because the game is played on an unknown tablet (iOS or Android) device using an unknown pair of headphones, the absolute level of sound from the headphones (or potentially earphones) is unknown. The calibration game is aimed at determining the softest speech sound the child can understand (expressed as the digital signal level sent to the tablet audio output). This level is compared to the softest speech sound that an adult *thought* to have no known hearing loss can understand, but the accuracy of the overall game, as will be explained later, is not dependent on the reference adult actually having good hearing. In the calibration game, five differently coloured birds or fish move across the tablet screen. First, the sounds are presented to both ears simultaneously for the adult. Following this, they are presented to one ear at a time for the child.   The player is instructed to tap the one whose colour matches the word (red, blue, black, white, yellow) that they hear. The sound level is varied adaptively, initially in 6 dB steps, then in 3 dB steps, to find the speech recognition threshold in quiet (SRTq), defined as the average of the levels presented after a practice phase has ceased. Testing concludes when the estimated standard error of the mean presentation level drops below 2 dB or when 15 sounds have been presented after practice. Calculation of the standard error takes into account the non-independence of the presentation SNRs (as each SNR can vary from the previous one by only the size of the step being used in the adaptive procedure).

The difference (ΔSRTq) between the poorer of the child's results for the two ears and the adult's speech reception thresholds provides one metric used in the final screening result, and the accuracy of this metric is of course dependent on the adult having good hearing. The presentation level for the competing sounds used in the subsequent tests (described in the next sections) is then set at 38 dB above the weighted average of the child's SRTq and the adult's SRTq. The child:adult weighting in the average is 80:20. This ratio was chosen so that while differences in headphone sensitivity are fully compensated, for children with hearing loss, the amplification will be less than the amount of their loss, as full compensation would produce excessive loudness for people with sensorineural loss.

Children with conductive loss or sensorineural loss are expected to have elevated SRTq. The difference in thresholds between the child's two ears also provides an estimate of inter-aural threshold difference, and is unaffected by the status of the adult's hearing or the nominal sensitivity of the headphones.

Speech-in-noise game

The speech-in-noise game requires the child to tap an object on the screen as each object's name is called out by the game characters. The objects form part of a story in which a park ranger sets out to locate an apparently lost ranger, and in so doing requires the help of the child being tested. To make the test sensitive to spatial processing disorder (Cameron & Dillon, 2008), two different competing sounds (each of which is ongoing speech from a single talker) are applied to the two ears using head-related transfer functions, one corresponding to a source at 90° to the right and one corresponding to a source at 90° to the left. The target signal is also applied to both ears, but using head-related transfer functions appropriate to a frontal sound. The level of the target sound, and hence the signal-to-noise ratio (SNR) is adjusted adaptively, increasing by 4 dB after an incorrect response, increasing by 2 dB after an absent response, and decreasing by 2 dB after a correct response. When no response is made, the item is repeated twice before moving onto the next item, so the number of items presented is increased for children who respond less frequently when uncertain.   If there are no failures to respond (just incorrect responses when uncertain), this combination of step sizes adapts to the 66% correct point on the child's psychometric function (Kaernbach, 1991). If there are no incorrect responses (just absent responses when uncertain), the test adapts to the 50% correct point on the psychometric function. Consequently, the target percentage correctly identified is somewhere between 50 and 66% depending on the caution of the

child when items are not clearly heard. The speech reception threshold in noise (SRTn) is the average of the SNRs presented after completion of an adaptive practice phase during which a larger step is used. Validity of the SRTn is evaluated by calculating the standard error of the presentation SNRs during the measurement phase, and deeming invalid any results for which the estimated standard error exceeds 2.5 dB. This value is the upper limit found during extensive pilot testing for the great majority of children, for whom the adaptive process results in a well-behaved adaptive track (i.e. the pattern of variation in SNR from trial to trial), showing a descending SNR during the practice phase followed by small fluctuations in SNR around the child's final SRTn during the measurement phase. Children with sensorineural loss and any auditory processing disorders that affect speech perception in noise are expected to perform outside normal limits on this test, but not children with conductive loss because the test is performed at a fixed sensation level, and they should not have cochlear distortion.

Tone-in-noise game

The tone-in-noise game is similar to the speech-in-noise game, except that the target sound is a pure tone, frequency modulated between 1400 Hz and 1600 Hz (representing a bird call), presented to both ears using a frontal head-related transfer function, and the competing sounds (applied to each ear using $\pm$90 HRTFs) are temporally modulated random noises, each of which has been band-stop filtered to have very little power between 1350 Hz and 1650 Hz. The two competing sounds (which in the story are "engine noises") differ both in their spectral shape below 1000 Hz and above 2000 Hz (which should not affect their masking ability) and in the rate at which they are temporally modulated. They thus have a noticeably different character in each ear. The child's task is to tap an on-screen button every time they hear a bird, for which they are rewarded with a "photo" of the bird. During the measurement phase the target level decreases by 3 dB when the sound is detected and increases by 4 dB when it is not detected. The test thus adapts to the 57% point on the tone detection psychometric function (Kaernbach, 1991).   The tone reception threshold in noise (TRTn) is the average of the SNRs presented after completion of an adaptive practice phase during which larger step sizes are used. As with the speech-in-noise test, the standard error during the measurement phase is calculated and tracks with excessive values are deemed invalid. In addition, tone-in-noise tests with excessive false button presses are deemed invalid.

Children with sensorineural loss are expected to perform outside normal limits on this test, because, unlike with conductive loss, sensorineural loss causes a widening of auditory filters, decreasing the child's ability to focus on the narrow frequency region containing the tone. People with sensorineural hearing loss also have reduced ability to detect sounds during the low-power portions of temporally modulated masking sounds (Bronkhurst & Plomp, 1992; Duquesnoy, 1983; Festen & Plomp, 1990).

Stimulus intelligibility equalisation

Items for the calibration (speech-in-quiet) game and for the speech-in-noise game were recorded as isolated words. Initially, the rms levels of their digital files were equalised and the game was played by 94 children and 30 adults with normal hearing. Throughout this paper, "normal hearing" refers to people screened to have hearing thresholds of 20 dB HL or better at 500, 1000, 2000 and 4000 Hz. The percentage of time each word in the speech-in-noise game was perceived correctly was calculated, and items with scores markedly lower than the rest were substituted by new items. The game was then played by a further 97 children (average age 5.5 years) with normal hearing and 41 children with hearing loss in at least one ear (average age 6.6 years). The four frequency average hearing loss (4FAHL; average of 500, 1000, 2000 and 4000 Hz) of the latter group in the poorer ear was 49.8 dB HL. The children with normal hearing were recruited from local schools and pre-schools, and those with hearing impairment were recruited from the caseload of Australian Hearing,

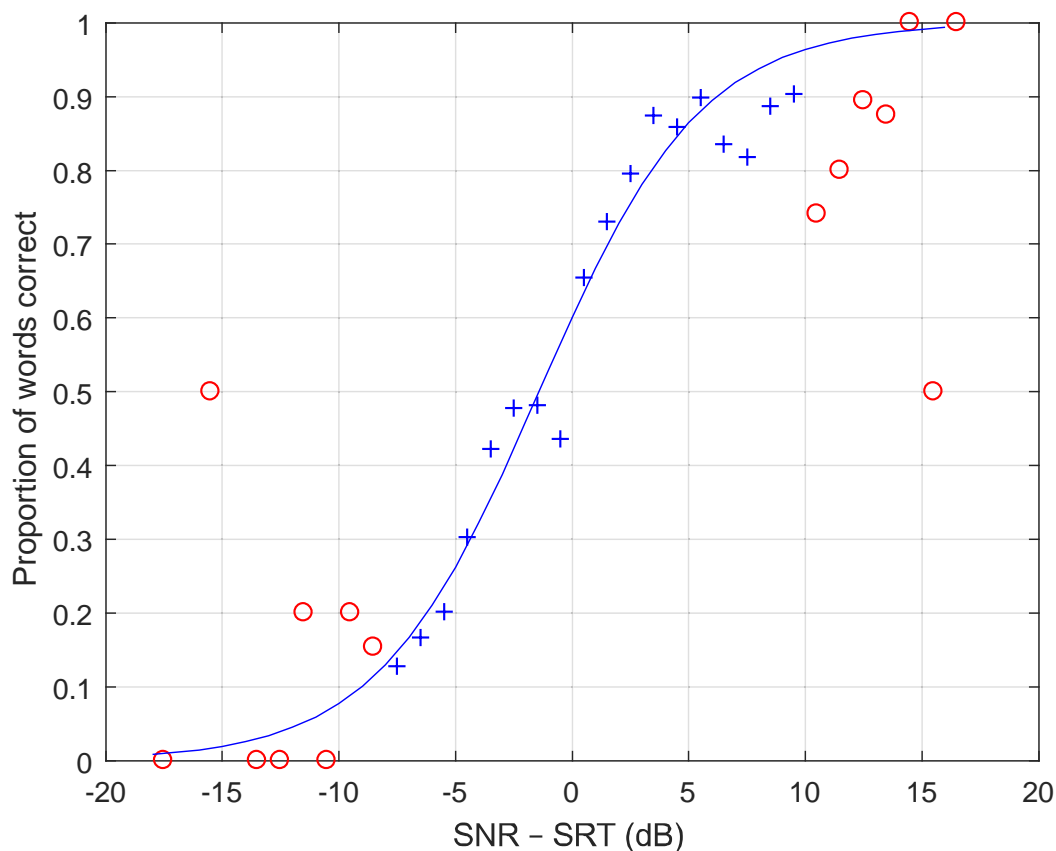Dillon: Sound Scouts hearing test game

from whom recent audiograms were also obtained. All testing reported in this paper was carried out using iPads. Although a variety of headphones were used, most of the measurements were made using Sennheiser HD215 headphones.

For each presentation of a target stimulus, the presentation level was expressed relative to the final SRTn measured for that respondent, and the correctness of the response noted. These relative presentation levels were grouped into bins 1 dB wide, and for each bin, the proportion of times the target sounds was correct was calculated. Initially, this was performed for all target stimuli combined, which resulted in 3707 observations from which the psychometric function shown in Figure 1 was formed. The logistic regression curve, as given in Equation 1, was fitted.

$$P = 1 / (1 + e^{(a + b.L)}) \quad .... 1$$

where P is the proportion correct, a and b are fitting constants, and L is the relative presentation level. The asymptotic slope of the fitted curve is 7.2% per dB. When the presentation level is equal to the final SRT (i.e. the relative presentation level is zero), the percentage correct is 60%, which is consistent with the value expected given the asymmetrical step sizes used.

Figure 1: Psychometric function (data and fitted function) for all stimuli combined, prior to stimulus level adjustment. Crosses show data points for which there were 30 or more responses per relative presentation level, and circles show the remaining (less reliable) data points.

To determine the relative difficulty for each of the 24 speech targets, the same process was repeated for each speech target individually. Because each fitted function is less accurately

Dillon: Sound Scouts hearing test game

estimated as it is based on only approximately one twenty-fourth of the data of the combined function, the *b* values were averaged across all the stimuli, and the fitting was repeated using this same *b* value, which results in an asymptotic slope of 8.2% per dB.

The SNR at which individual curves passed through the 60% correct point varied from -4.0 dB to 5.5 dB, which is a total range of 9.5 dB in variation of item difficulty. The level of each item was adjusted by the difference between the SNR at each item's 60% point and the average (which was -0.1 dB) of the individual SNRs for 60%. All adjustments were less than 6 dB, and all but three were less than 3 dB. This intelligibility equalisation maximises the accuracy of adaptive speech tests by minimising the variation in SNR as the test progresses (Dillon, 1982).

Study 2: Normative data and z score calculation

Methods

The game (with equi-difficulty stimuli) was played by 213 children and 20 adults screened to have hearing thresholds of 20 dB or better from 500 Hz to 4000 Hz. The children's ages ranged from 4.1 to 15 years, with a mean of 7.8 years. (Due to a limitation in the test software at the time of data collection, children older than 14.0 years had their age recorded in the software as 14.0 years, and were analysed as such.) The adult's ages ranged from 18 to 53, with a mean of 34.3 years. Children were recruited from local schools or pre-schools and were supervised by one of seven adults (comprising audiologists, research scientists, and test developers), all of whom also had normal hearing, who performed the adult calibration task independently for each child. Testing was carried out in the quietest room available in each school or pre-school. Results include both sets of scores for 43 children who were re-tested one month later to determine test-retest differences, as the mean test-retest differences were small. The normative data were used to derive a method for calculating z-scores, in a manner that enabled the normal variation of performance with age to be allowed for.

The most reliable overall test metric (which will be referred to as the H metric) for the detection of sensorineural hearing loss is, in principle, produced by combining the three z-scores (one for each test), because sensorineural loss is expected to affect all three test scores. Combination in z-score units has the advantage that the three tests are then inherently weighted in a way that reflects the spread of results in the normal hearing population, including any differences in the precision of each test. The scores were combined using the squares of each z-score (when they were negative; positive z-scores were discarded in the sum), to give greatest weight to whichever individual metric showed the greatest hearing deficit. This ensured that negative scores on one test could not be offset by positive scores on another test. Speech-in-noise z scores and tone-in-noise z scores were each excluded from the composite H metric whenever their validity check (described earlier) indicated that the test result could not
be relied on. (In the rare case where both of these test results are invalid, Sound Scouts reports that the test results are invalid.)

To set cut-off values for detection of hearing loss, the H metric was calculated from the results of 116 children comprising 60 children with normal hearing (mean age 8.3 years), 37 children with sensorineural hearing loss (mean age 9.5 years), and 19 children with conductive loss (mean age 9.4 years) at the time of testing. For those children tested more than once, results from only the first administration of the test to each child were included in the data. Children with normal hearing were recruited from local schools and pre-schools. Children with sensorineural loss were recruited from Australian Hearing clinics on the basis of availability. Those with conductive loss were recruited either from several audiology clinics on the basis of availability or from local schools when

conventional audiometry had shown that children being screened for normal hearing actually had a conductve loss in at least one ear.

Table 1 shows the criteria that were adopted to determine whether a particular test result could be caused by any of a sensorineural hearing loss, a conductive loss, or an auditory processing disorder (APD). The logic of each constraint was based on a theoretical expectation of how each type of loss *should* affect each subtest. The actual boundary criterion against which each score was compared was partly influenced by the data obtained from administering the test to the same 116 children described earlier in this section with known hearing status (normal, sensorineural or conductive).

--- Table 1 about here ---

For one of the potential causes to be accepted as a possible cause, all three requirements in the corresponding row must be met. Each condition had to be met only when the corresponding test score passed the validity criteria for that score.

Results

Calibration test

The extent to which the poorer of the SRTq of the two ears for the player exceeded the binaural SRTq of the supervising adult (i.e ΔSRTq) is shown as a function of age in Figure 2. Deviations from the regression curve had a standard deviation of 4.5 dB for the children. Consequently, the speech-in-noise score for any child can be expressed as a z score, with the effect of age allowed for, by Equation 2. More negative z scores indicate that, relative to a supervising adult with good hearing, a child needs a presentation level in quiet higher than is typical for the child's age-matched peers.

$$\Delta SRTq \text{ z score} = -(\Delta SRTq - 1.1 - 28.9 * e^{-Age/3.94})/4.54 \qquad ..... 2$$
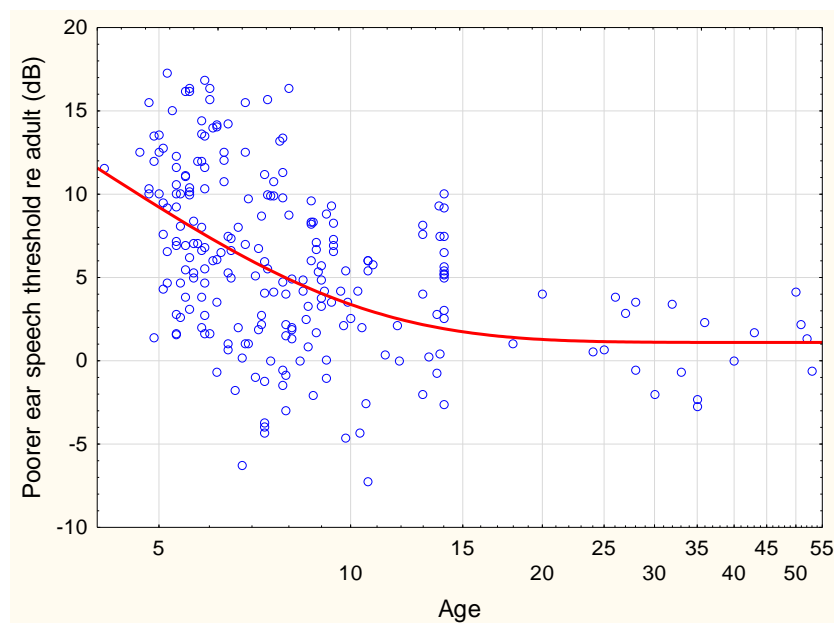

Figure 2: Variation of ΔSRTq with age and the fitted exponential regression.

Speech-in-noise test

The speech-in-noise data (excluding those that had standard errors exceeding the acceptance criterion) are shown versus age in Figure 3 for the 205 children and 19 adults for whom the adaptive tracks met the validity criteria. Deviations from the regression curve had a standard deviation of 2.2

Dillon: Sound Scouts hearing test game

dB for the children. Consequently, the speech-in-noise score for any child can be expressed as a z score, with the effect of age allowed for, by Equation 3.

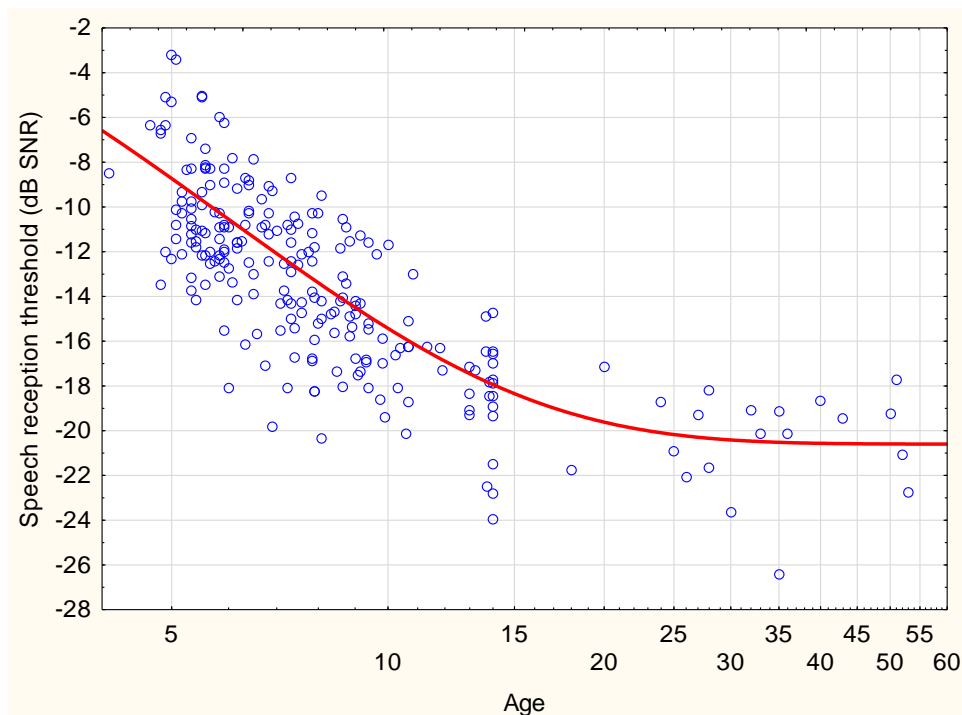SRTn z score = -(SRTn + 20.6 - 27.24*e$^{-Age/6.02}$)/2.20   ….. 3

Figure 3: Variation of speech reception threshold in noise (SRTn) with age and the fitted exponential regression.

Tone-in-noise test

Tone reception thresholds are plotted as a function of age in Figure 4 for the 156 children and 19 adults for whom the tone-in-noise results met the validity check. An exponential function of age was fit to the data, excluding the four children shown who obtained scores much better than the other 152 children. Deviations from the regression curve had a standard deviation of 3.1 dB for the children. Consequently, the speech-in-noise score for any child can be expressed as a z score, with the small (and not significant) effect of age allowed for, by Equation 4.

TRTn z score = -(TRTn + 40.7 − 41.0*e$^{-Age/1.42}$)/3.14   ….. 4

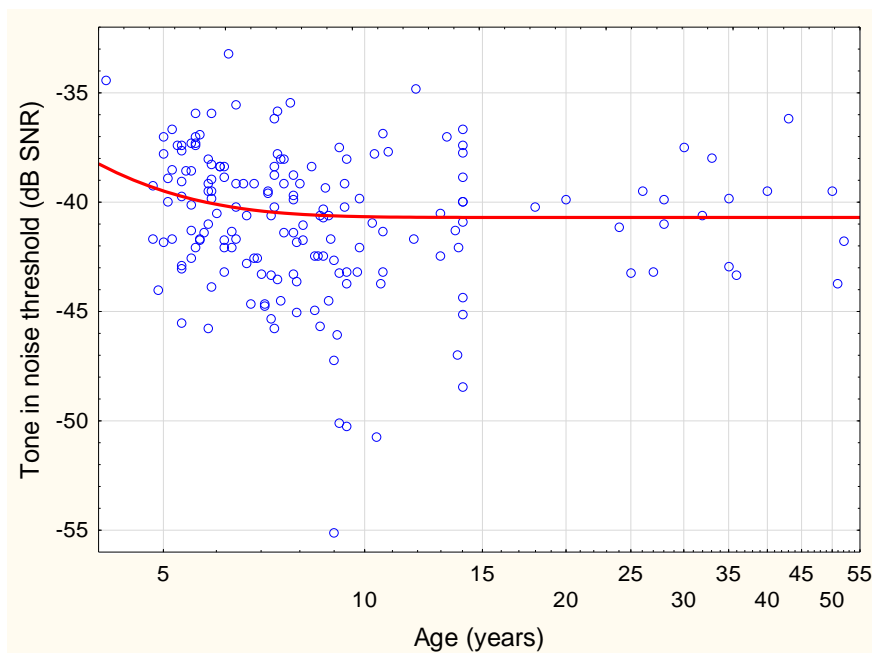Dillon: Sound Scouts hearing test game

Figure 4: Variation of tone reception threshold in noise (TRTn) with age and the fitted exponential regression.

Overall test metric

Figure 5 shows the overall test metric, H, as a function of 4FAHL in the poorer ear for 108 of the 116 children with known hearing status. The test results were incomplete (due to the adaptive tracks not meeting the validity criteria) for 8 children (3 normal hearing, 3 sensorineural and 3 conductive), and these results are not included in the graph. Based on these results, criteria were established for classifying the results into "pass", "fail" and "inconclusive" with boundaries for the H metric at -1.8 and -2.5. The latter region, which in the most recent version of Sound Scouts is reported as "borderline", is intermediate to the "pass" and "fail" regions, and we considered it unsafe to report a definite classification of results in this range. (The game reports children obtaining a score in the "fail" region with a description more nuanced than using the term "fail".)
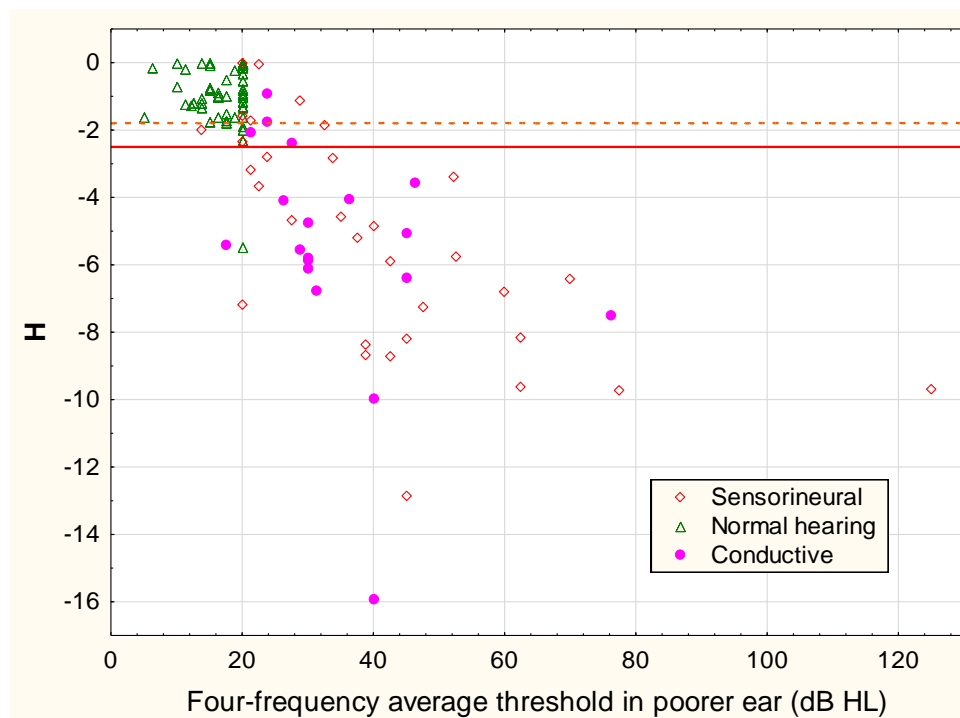
Figure 5: Composite test metric, H, versus 4FAHL in the poorer ear for children with known hearing status as indicated in the legend. The dashed line shows the boundary established between "pass" and "inconclusive" and the solid line shows the boundary between "inconclusive" and "fail".

Examination of Figure 5 shows that, of the children with either type of hearing loss, the largest hearing loss in the poorer ear that enabled a "pass" was 29 dB 4FAHL. This child had 4FAHL thresholds of only 6 dB in the better ear. The results from the same 116 children (including those with an incomplete result) are categorized in Table 2.

--- Table 2 about here ---

Of the 98 children for whom Sound Scouts indicated a definite result (i.e. pass or fail), 90 received the correct result, and the remaining 8 received the incorrect result. Of these 8, one child with normal hearing had an overall H metric score far outside the normal region, caused by the speech-in-noise test score giving a very negative z score. The cause is unknown, but could represent an auditory processing disorder previously unrecognised in this child. Of the 7 children with hearing loss of some sort who received a pass result, all had 4FAHL less than 30 dB HL in the poorer ear and less than 23 dB HL in the better ear.

To calculate sensitivity and specificity, a decision must be made about how to treat both the incomplete and inconclusive Sound Scouts screening results. In principle, they can be treated either as a pass, a fail, or excluded from the analysis. The game advises that for these children, the test should be repeated. Sensitivity and specificity results are shown in Table 3, with incomplete results excluded from the analysis and inconclusive results alternative treated as though they are pass, fail, or excluded. The latter is appropriate if all children obtaining an inconclusive result repeated the test and obtained a more definite result on the retest.

--- Table 3 about here ---

Accuracy of inferred probable reason for fail

Because Sound Scouts has been designed to report each of the causes of a fail or inconclusive result that seems consistent with the pattern of results obtained, application of the criteria in Table 1 to the 40 children with a fail result identified two potential causes for 19 children, one potential cause

for 16 children and no potential cause for the remaining 5 children. Of those 16 cases where a single cause was inferred, the correct cause was inferred for only 10 of them.

Discussion

Our aims were to produce a game that was engaging and fun for children to play, detected hearing problems, and differentiated sensorineural loss, conductive loss and auditory processing disorders from each other. Based on many anecdotal comments from children and observation of the children as they played the game, we are confident we have achieved the first of these aims, though we have not sought to empirically validate this.

The results in this paper show that the second aim was achieved. The combination of the age-adjusted scores from the three tests, one based on speech in quiet, one based on speech in noise, and one based on tone in noise, produced a composite metric that reliably detected children with a 4FA hearing loss greater than 30 dB HL in either ear. Although we did not show how each test individually detected hearing loss, deriving the overall result from a combination of the three scores had a better combination of sensitivity and specificity than relying on any one of the tests in isolation. The improvement in sensitivity/specificity when test results are averaged, albeit in a non-linear manner, is not surprising as averaging diminishes the impact of random measurement error in each of the measures. Although the sensitivity was "only" 85% (with inconclusive results excluded) for detecting any hearing loss outside the normal range, if the aim were to detect children with 4FA hearing thresholds greater than 30 dB HL in the poorer ear, the sensitivity was actually 100%, combined with a specificity of 98%. In practice, we would expect a specificity lower than this, if children are tested who are not proficient in English because it is their second language.

The third aim, however, was not achieved. The differentiation of conductive from sensorineural loss, even in those cases where Sound Scouts inferred a single potential cause, was correct only two-thirds of the time. It is worth speculating on why our hypothesised basis for distinguishing between conductive and sensorineural losses did not work as well as expected. In brief, we expected that children with conductive loss would perform as well in noise as children with normal hearing once the stimuli had been amplified to compensate for the attenuating effects of the conductive loss. One reason why this might not have worked is because the amplification applied was less than the degree of hearing loss, a precaution needed because full amplification would likely provide an uncomfortable loudness experience for those children who actually had sensorineural hearing loss. Second, because we do not measure the hearing thresholds of the children as a function of frequency, the amplification is provided equally at all frequencies, so a normal loudness balance will usually not be achieved. Third, like all measurements, each test score has a random error component, which restricts the accuracy with which different scores can be compared. Fourth, it is possible that children who have had a conductive hearing loss for weeks, months, or longer, do not immediately "hear normally" even if amplification reverses the attenuation caused by the middle ear. This would be the case if their brain has changed the way that it processes sounds to compensate for the attenuation caused by the conductive loss (Brotherton et al. (2016, 2017). Fifth, if conductive loss has been present for a long time, it may have impacted on their language acquisition sufficiently to make in noise recognition worse than their age peers, just as can occur for sensorineural loss. Finally, and possibly most importantly, we suspect that the very negative SNRs at which the tone is normally detected has resulted in the tone threshold sometimes being effectively an absolute threshold, rather than a masked threshold. Further research into this is underway and we anticipate that if the problem can be circumvented with a modified target and/or masker, the tone-in-noise test will become a more effective part of the test battery, potentially better differentiating the different types of hearing problems.

Dillon: Sound Scouts hearing test game

Prior research has indicated that five-year-old children need a SNR 3 to 5 dB better than adults to perceive speech equally well in noise (Boothroyd, 1997; Gravel et al, 1999). These data add to that by showing SRTn as a continuous function of age. From age 5 to 10 years, each additional year of age results, on average, in an improvement in SRTn by 1.2 dB (Figure 2). Performance at age 10 was still 4 dB short of what the adults were capable of. A likely explanation for the greater age effect in this experiment is that the target speech and distracting speech are spatially separated in Sound Scouts (and usually also in real life), and the ability to use spatial cues to segregate target speech from competing speech also improves markedly over this same age range (Cameron & Dillon, 2007). By contrast, the improvement of tone detection in noise with age was negligible. This is likely because increasing familiarity with language is not an issue, but it is also possible that children have less damaged cochlea, and hence more sharply tuned auditory filters than adults, even adults with hearing thresholds in the normal range. Although the tone and noise stimuli were also spatially separated, their strongly dissimilar sound makes it less necessary to use spatial cues to segregate the two sounds.

We have not yet begun to evaluate the accuracy with which Sound Scouts can identify an auditory processing disorder. This is a daunting task, given the wide variety of deficits that probably fall under the broad term "auditory processing disorder", the difficulty in unambiguously diagnosing many of these deficits, and the minimum age restrictions (typically seven years of age or greater) that apply to tests of auditory processing disorder. Sound Scouts is able to quantify whether the SNR needed by, say, a five-year old child to understand 60% of speech items in a noise background is less than or greater than the SNR needed by an average five-year-old, and the extent of the deviation from average. Greater than average difficulty in noise could be caused by an auditory processing disorder (of whatever type), a language deficit (including being tested in a child's second language), or a cognitive deficit (such as in working memory, attention, or general fluid intelligence). Whatever the reason for needing a higher SNR than age peers, in difficult listening situations, the child will either understand less than his or her age peers, or will have to devote a greater portion of his or her cognitive capacity to filling in the words not understood, leaving less available for learning. The measure of speech-in-noise ability therefore seems like important information to have about a child for whom there is some doubt about their auditory functioning, even if there is no existing test of auditory processing disorder against which the Sound Scouts results can be compared. In future research we hope to compare the Sound Scouts scores against scores determined from parent-report questionnaires about listening behaviour, and so provide some cross-validity for Sound Scouts in the younger age range.

In brief the inclusion of three different hearing tests within the game proved to be valuable because of the redundancy and averaging implicit in having multiple tests, not because of the ability to compare results to ascertain the type of hearing problem. More sophisticated algorithms for determining a threshold from adaptive tracks that display a lot of variability have since been developed. Further evaluation will be needed to evaluate whether these changes, plus the altered tonal stimuli described earlier, will enable the type of hearing problem to be more accurately determined.

Limitations

The evaluation of accuracy was carried out using the same data used to determine the cut-off values that determine "pass" and "fail" results. It is probable that the sensitivity or specificity with which the game differentiates children with normal hearing from those with hearing loss will be less if a new group of children were to be tested, or if children were to be tested under different conditions (e.g. unsupervised, or in noisier conditions). Also, in real life the adult performing the initial calibration thresholds will have unknown hearing status, which in some cases will affect the results of the speech-in-quiet test for the child, but not the remaining two tests. Conversely, more

Dillon: Sound Scouts hearing test game

sophisticated tests of the internal reliability of the adaptive tracks and calculation of the threshold represented by each track have been added since the results reported in this paper, which should enhance accuracy. The results reported here were all obtained using iPad devices. The results reported here were all obtained using iPad devices. Subsequent objective testing of the sound output has shown that, as expected, the sounds output by the device are not affected by whether it is an iPad or Android device.

Other potential applications

Although the original motivation for the test was to detect previously undiagnosed hearing problems in children around school entry age, or older, the method has other potential applications.

Post-treatment monitoring

Following detection of significant conductive loss, there is a need to determine whether treatment (antibiotics or ear ventilation tubes) or watchful waiting have resulted in correction of the hearing loss. This is usually achieved by an audiologist assessing hearing by tympanometry or hearing thresholds, or by a medical practitioner otoscopically examining the eardrums. Both of these involve significant expense to the health system and/or to parents. On-line tests like Sound Scouts can provide a lower cost, more convenient method by which hearing status after treatment can be monitored.

In addition, it is known that protracted otitis media in early childhood strongly pre-disposes children to acquire spatial processing disorder (Cameron et al., 2014; Tomlin & Rance, 2014). Although the sensitivity of Sound Scouts to spatial processing disorder has not yet been evaluated, the test was designed to be sensitive to this condition, so its use after treatment for otitis media may also detect those children for whom otitis media has resulted in spatial processing disorder, for which separate remediation is needed, is already available, and is effective (Cameron et al., 2012). Additional stories are being developed so that the score on re-testing is not affected by familiarity with the items in the story. The test can be used for this purpose for any children over approximately 4.5 years of age, which is lower than the minimum of 6 years recommended for the LiSN-S test of spatial processing disorders (Cameron & Dillon, 2007).

Acknowledgements

Declaration of Interest

Sound Scouts Pty Ltd and the National Acoustic Laboratories receive revenue when Sound Scouts is downloaded and used.

References

Australian Hearing (2016). Aided child demographics. https://www.hearing.com.au/wp-content/uploads/2016/12/Aided-Child-Demographics-Dec-2016.pdf

Barry JG, Ferguson MA, & Moore DR. 2010. Making sense of listening: the IMAP test battery. *J Vis Exp* (44).

Dillon: Sound Scouts hearing test game

Bess F.H., Dodd-Murphy J., Parker R.A. 1998. Children with minimal sensorineural hearing loss: prevalence, educational performance, and functional status. *Ear Hear*, 19(5), 339-354.

Boothroyd A. 1997. Auditory development of the hearing child. *Scand Audiol Suppl,* 46, 9-16.

Bronkhorst A.W. & Plomp R. 1992. Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *J Acoust Soc Am*, 92(6), 3132-3139.

Brotherton H., Plack C.J., Schaette R., & Munro K.J. 2016. Time course and frequency specificity of sub-cortical plasticity in adults following acute unilateral deprivation. *Hear Res*, 341, 210-219.

Brotherton H., Plack C.J., Schaette R., & Munro K.J. 2017. Using acoustic reflex threshold, auditory brainstem response and loudness judgments to investigate changes in neural gain following acute unilateral deprivation in normal hearing adults. *Hear Res*, 345, 88-95.

Cameron S. & Dillon H. 2007. Development of the Listening in Spatialized Noise – Sentences Test (LISN-S). *Ear Hear*, 28(2), 196-211.

Cameron S. & Dillon H. 2008. The listening in spatialized noise-sentences test (LiSN-S): comparison to the prototype LiSN and results from children with either a suspected (central) auditory processing disorder or a confirmed language disorder. *J Am Acad Audiol,* 19(5), 377-391.

Cameron S., Dillon H., Glyde H., Kanthan S. & Kania A. 2014. Prevalence and remediation of spatial processing disorder (SPD) in Indigenous children in regional Australia. *Int J Audiol,* 53(5), 326-335.

Cameron S., Glyde H., & Dillon H. 2012. Efficacy of the LiSN & Learn auditory training software: Randomized blinded control study. *Audiology Research*, 2(1), 86-93.

Dillon H. 1982. A quantitative examination of the sources of speech discrimination test score variability. *Ear Hear,* 3(2), 51-58.

Duquesnoy A.J. 1983. Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *J Acoust Soc Amer,* 74, 739-743.

Festen J. & Plomp R. 1990. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Amer,* 88(4), 1725-1736.

Gravel J.S., Fausel N., Liskow C., & Chobot J. 1999. Children's speech recognition in noise using omni-directional and dual- microphone hearing aid technology. *Ear & Hear,* 20(1), 1-11.

Kaernbach, C. 1991. Simple adaptive testing with the weighted up-down method. *Attention, Perception, & Psychophysics*, 49(3): 227-229.

Kuppler K., Lewis M. & Evans A.K. 2013. A review of unilateral hearing loss and academic performance: is it time to reassess traditional dogmata? *Int J Pediatr Otorhinolaryngol,* 77(5), 617-622.

Nillson D., Dillon H., Gardner-Berry K., & King A. In preparation. Why do so many children get their first hearing aids during the early school years?

Tomlin D. & Rance G. 2014. Long-term hearing deficits after childhood middle ear disease. *Ear Hear,* 35(6), e233-242.

Yeung J., Javidnia H., Heley S., Beauregard Y., Champagne S., & Bromwich M. 2013. The new age of play audiometry: prospective validation testing of an iPad-based play audiometer. *J Otolaryngol Head Neck Surg,* 42, 21.

Dillon: Sound Scouts hearing test game

Table 1. Criteria used to determine if each type of hearing problem could cause the pattern of results observed.

	Calibration game	*Tone in noise*	*Speech in noise*
Sensorineural loss	$\Delta SRTq_z < -2$ or asymmetrical loss detected	$TRTn_z < -1$ or asymmetrical loss detected	$SRTn_z < -1$
Conductive loss	($\Delta SRTq_z < -2$ and $\Delta SRTq_z > -6.5$) or asymmetrical loss detected	$TRTn_z > -3$	$SRTn_z > -4$
APD	$\Delta SRTq_z > -2$ and not asymmetrical	No requirement	$SRTn_z < -2$

Table 2. Sound Scouts screening results for 116 children with known hearing status.

	Overall Sound Scouts result				
	Pass	**Inconclusive**	**Fail**	**Incomplete**	**Total**
Normal hearing	51	3	1	5	**60**
Sensorineural loss	5	3	24	5	**37**
Conductive loss	2	2	15	0	**19**
Total	**58**	**8**	**40**	**10**	**116**

Table 3. Sensitivity and specificity of the Sound Scouts screening result with inconclusive results treated in different ways.

	Inconclusive results treated as pass	**Inconclusive results treated as fail**	**Inconclusive results excluded**
Sensitivity	76%	86%	85%
Specificity	98%	93%	98%