# An opinion on the assessment of people who may have an auditory processing disorder

Harvey Dillon[1,2], Sharon Cameron[1,2], Helen Glyde[1,2], Wayne Wilson[3], Dani Tomlin[2,4]

[1] National Acoustic Laboratories

[2] The Hearing Cooperative Research Centre

[3] University of Queensland

[4] University of Melbourne

Revised: 11 May 2011

**Address for correspondence:**

Dr Harvey Dillon
NAL
126 Greville St
Chatswood, 2067
Australia
Harvey.dillon@nal.gov.au
Phone: +61 408 067 682

**Abstract**

We need to rethink how we assess auditory processing disorder (APD). The current use of test batteries, while necessary and well-accepted, is at risk of failing as the size of these batteries increases. To counter the statistical, fatigue and clinical efficiency problems of large test batteries, we propose a hierarchical approach to APD assessment. This begins with an overall test of listening difficulty in which performance is measurably affected for anyone with an impaired ability to understand speech in difficult listening conditions. It proceeds with a master test battery containing a small number of single tests, each of which assesses a different group of skills necessary for understanding speech in difficult listening conditions. It ends with a detailed test battery, where the individual tests administered from this battery are only those that differentiate the skills assessed by the failed test/s from the master test battery, so that the specific form of APD can be diagnosed. An example of how hierarchical interpretation of test results could be performed is illustrated using the Listening in Spatialized Noise Sentences test (LiSN-S). Although consideration of what abilities fall within the realm of auditory processing should remain an important issue for research, we argue that patients will be best served by focusing on whether they have difficulty understanding speech, identifying the specific characteristics of this difficulty, and specifically remediating and/or managing those characteristics.

## Introduction

Guidelines for testing auditory processing disorder (APD) recommend the use of a test battery, with the tests in the battery carefully chosen so that known or hypothesized disorders are likely to produce aberrant test scores on one or more of the tests (ASHA, 2005; AAA, 2010; BSA, 2010). Though the ASHA guidelines call for the tests included in the battery to be based on the presenting symptoms, the body of knowledge needed to appropriately choose these tests does not exist. In practice, each clinician adopts a standard battery (with the particular set of tests varying from clinician to clinician and from country to country). When results on some predetermined number of the tests are aberrant by more than some predetermined amount, the patient is diagnosed as "having APD".

In this article, we discuss the difficulties that arise from the current conceptualization of APD and its assessment, and we present an alternate hierarchical approach to the APD test battery. We argue that the focus should instead be on identifying patients who have more than usual difficulty understanding speech in (real-life) difficult listening conditions, and then applying differential tests in a structured and results-dependent manner to enable the cause and magnitude of the difficulty to be determined, and preferably remediated. Our research experience has primarily been focused on children reported as having difficulty understanding speech in the classroom, despite having normal hearing sensitivity. The opinions in this article reflect that focus, though we expect the issues discussed will be relevant to other populations who report listening difficulties not commensurate with their hearing thresholds.

In presenting an alternate hierarchical approach to the APD test battery, we note that listening difficulties can be wide and varied and may include difficulty following instruction, comprehending speech, or attending to auditory stimuli. These difficulties may occur in noise or quiet, reverberant conditions or with rapid speech. Furthermore, listening difficulties may be the result of many disorders including disorders of perceptual processing, sensory attention or higher level cognition. As a result, we deliberately do not attempt to define APD. Rather, we choose to focus on the diagnosis and management of listening difficulties. Despite several consensus and position statements, disagreement abounds over what constitutes an APD. These issues may well be just matters of arbitrary definition, incapable of resolution by any scientific experiment, though in the fullness of time, some definitions will likely be found to be more helpful than others. We conclude our article with further discussion on this issue.

**Problems with basing APD testing on a fixed test battery**

Humans can discriminate complex sounds along a large variety of dimensions. This applies to both speech and non-speech sounds. For any discrimination ability postulated to be important, tests of this discrimination can be created; often several different versions exist with different hypothesized advantages and disadvantages. Many such tests have been created; many more tests could, should, and probably will, be created.

For most of the discrimination abilities tested, it is conceivable that the ability is used during the perception of speech, either for segregating a target signal from other sounds, or for analyzing and recognizing the sounds of speech. The resulting discriminations may be essential for understanding speech if the speech signal is not sufficiently redundant, or if redundant information is unavailable because of noise, reverberation, or some other factor.

It is sometimes argued that the only tests that can be validly used to infer the presence of an APD are those with known links to a physical lesion in a particular part of the brain (American Academy of Audiology, 2010; American Speech-Language-Hearing Association, 2005; Musiek et al, 2005; Baran, 2007). Requiring such a link appears to be an unnecessary restriction on the development of useful tests. There can be no doubt that neurons are required to carry out any auditory task, and that a lesion at the sites of those neurons would impair performance. This is true whether or not those sites have as yet been discovered through imaging studies or surgery involving patients who have physically identifiable lesions. Furthermore, it is not the case that each auditory discrimination relies on only one brain site or that each brain site facilitates only one discrimination ability. Finally, it is certainly possible to calculate, the sensitivity and specificity with which an APD test can detect the presence of a lesion that is observable in a brain scan, as the brain scan provides the gold standard that differentiates the person from the general population (Hurley and Musiek, 1997). There is no reason to suppose, however, that the same test sensitivity and specificity would apply to a child who had a developmental delay caused by partial but inadequate development of neural connections in the same region of the brain. In all fields, the greater the difference between the characteristic in the affected population and the characteristic in the regular population, the more favorable the sensitivity/specificity combination of any test. Tests cannot therefore have an inherent sensitivity and specificity that is applicable irrespective of the magnitude of the disorder. In the case of a person suspected to have some form of CAPD, there is not even any way to calculate the sensitivity and specificity of the test for that CAPD, as there is no absolute gold standard against which the test results can be compared. None of this argues against the usefulness of lesion studies. These studies (e.g. Bamiou et al, 2007; Musiek and Sachs, 1980; Musiek et al, 1980; Musiek et al, 1984; Musiek et al, 1985) are clearly an important mechanism for developing our knowledge about what processing occurs where in the brain (though sometimes the "sites" are very broad, such as somewhere in the vicinity of the Sylvian fissure [Boscariol et al, 2010], an area where most auditory processing is believed to take place), and for suggesting probable physical interpretations of distinctive test result patterns. Rather, we are arguing against *requiring* a site of lesion be known before a test can be accepted as indicating the proficiency of a child on a particular auditory task.

As stated previously, the existing approach to behavioral APD testing is to administer a standard battery of tasks, and set some criteria by which a pass or a fail is determined for each test, as well as a criterion by which an overall pass or fail is determined. A currently recommended criterion for diagnosing APD requires scores of two standard deviations or more below the mean for at least one ear on at least two different behavioral tests of APD (American Academy of Audiology, 2010). However, opinions about which tests should be in the standard battery vary from country to country, laboratory to laboratory, and time to time, and the number of tests in these batteries is increasing. As the number of tests increases, so does the likelihood that a person will fail at least one test, due to both accumulation of statistical probability and the fatigue that is likely to result from a long test battery. Two solutions to this problem have been proposed, both with severe disadvantages:

1. *Tighten the pass-fail criterion on each test*. (e.g. set the cut-off at three standard deviations (S.D.s) instead of two S.D.s. This is the equivalent of applying a

Bonferonni correction in multiple hypothesis testing. Unfortunately, this correction means that an individual has to be very aberrant in performance before a fail is recorded. The greater the number of tests, the tighter the criterion would have to be, and the more aberrant an individual has to be before he or she fails any one test, which is counterintuitive. Whether or not an individual has a specific abnormality should not, in principle, be affected by the number of tests in the battery.

2. *Require that the individual fail more than one test in the battery*. This is sensible only if it is known that every specific disorder of interest causes an aberrant performance on more than one test in the battery. This could be achieved by designing the tests to be deliberately redundant. Consequently, the number of tests must be greater, perhaps much greater, than the number of specific disorders being tested for. This further increases the statistical and fatigue problems caused by having multiple tests. Furthermore, in current batteries, when abnormal results are obtained on two tests, there is often no parsimonious reason why sub-normal functioning of any single neural processing structure (e.g. interhemispheric transfer via the corpus callosum, or processing based on inter-aural time or level differences) would cause a fail on those two particular tests, and a pass on the remaining tests. For example, how does one interpret a failed frequency pattern test combined with an abnormally low binaural masking level difference? Such a result would require reduced ability at multiple levels in the central auditory nervous system, and while such failure may occur, there is no reason to believe that all children who have listening difficulties have such widespread failures of auditory processing. Nonetheless, when a fail is recorded on the two tests, an overall diagnosis of APD is given, but no insight is developed

as to what might cause failure on two tests addressing apparently unconnected auditory abilities. In short, requiring a fail on more than one test is logical and efficient in test time only if APD were to be a generalized disorder that always caused sub-normal performance on a range of auditory abilities. We have established that some children experiencing listening difficulties in real life appear to have a single specific deficit: an under-developed ability to suppress sounds coming from directions different from the target, but otherwise appear to process sounds normally (Cameron and Dillon, 2008). This normal processing even extends to the perception of speech in competing signals, provided the target and the competing signals come from the same direction.

A third solution that could be proposed is to repeat any test that produces a failed result, and accept the result as genuine only if a fail occurs on both occasions. This approach introduces additional complexities if, on average, retest scores are higher than for the original test. The major problem, however, is that repeat testing is an inefficient use of test time.

A more fundamental problem with determining pass/fail criteria, even for a single test used in isolation, and even when each test has a known distribution of scores for people with no disorder, is that it is largely unknown how large a deficit is needed before the patient actually has a problem in real life. That is, what level of deficit on the test is necessary before the underlying ability is sufficiently aberrant to cause decreased ability to understand speech in commonly encountered listening conditions? This highlights the need for an objective measure of the degree of listening difficulty experienced. The level of deficit in the ability that is needed to create problems in real life may be much larger than the pass-fail criterion determined solely on the basis of a certain number of standard deviations of people who have not reported marked difficulties in understanding speech in difficult listening conditions. A fail does not necessarily indicate that the patient actually has a problem! Conversely, even people on the lower tail of some auditory skill, but still within the usual two-standard deviation limit, may experience more difficulty listening in noise than people with average ability on the auditory skill tested. An ingredient of APD research that has largely been missing is therefore a determination of *how* deficient any auditory ability (as assessed by a score on some test) has to be before it is associated with some specified level of increased difficulties in real life. Only a small amount of information on this is available from case studies (e.g. Musiek et al, 1985). For people with sensorineural hearing loss, the importance of different degrees of impairment (assessed on the basis of the audiogram) is assessed from the perspective of the disability (also known as activity limitation) that the impairment causes and from the handicap (also known as participation restriction) that then results. The parallel considerations in people with APD have not yet adequately been investigated, at least to the point of systematically determining test pass/fail criteria, including in our own research related to one specific form of APD. Of course, a paucity of research data linking scores on assessment tests to the magnitude of real-life problems does not preclude clinicians from routinely obtaining reports from a patient (or parent or teacher) to estimate the real-life effects that might be the results of any deficits revealed by performance tests.

To illustrate, let's concoct an example outside the field of auditory processing: Suppose we measure the ratio of index finger length to thumb length for children of a certain age, compute the mean and standard deviation, and hence calculate cut-off scores against which we can assess whether the child has "finger-thumb ratio disorder". If the ratio has a normal distribution across children, we could confidently predict that 5% of the children will be more than 2 S.D.s away from the mean, and half of these will be below the mean by more than 2 S.D.s. If the distribution has tails more extended than occurs for a normal distribution, then the proportion with this disorder will be even higher. Are any activities in their lives disadvantaged by having this disorder? Possibly yes, particularly in the extreme case of either digit being completely missing, but there is no reason why the ratio at which real-life problems emerge is at a value (e.g. 2 SDs below the mean) chosen statistically just so that the proportion of children who fail the test is manageably small.

It therefore seems very problematic to base a decision about whether an APD is present on the result of a pre-defined test battery, with each element of the test battery designed to be sensitive to a distinct processing domain (but actually relying on a range of abilities to perform the test), and with pass-fail boundaries determined solely on the basis of normative data. As more and more tests are invented, the difficulties of this approach will become increasingly apparent.

Should we invent more tests to detect forms of APD that have not yet been discovered? We think so. When children with self-reported, or teacher-reported hearing problems in the classroom, but who were within normal limits on the Weschler Intelligence Scale for Children, were assessed with a battery of commonly used tests, and also with a novel test of spatial processing ability - the Listening in Spatialized Noise Sentences test (LiSN-S; Cameron and Dillon, 2009) - most of the children were within normal limits on all the established tests used (random gap test, dichotic digits, masking level difference, pitch pattern sequence), were within normal limits in the LiSN-S conditions where speech and noise came from the same direction, but were outside normal limits on the LiSN-S whenever the target and distracters came from different directions (Cameron and Dillon, 2008). Until this test was devised, the presence of this specific form of APD in these children would have remained undiagnosed. One wonders how many further specific types of auditory processing disorder are waiting to be discovered, each of which will require one or more tests to diagnose.

**Alternative conceptualization of testing**

An alternative approach to assessing APD, in which the test battery is hierarchically structured, is shown in Figure 1. The model is adaptive depending on previous test results and the test battery is administered only when certain pre-conditions are met.

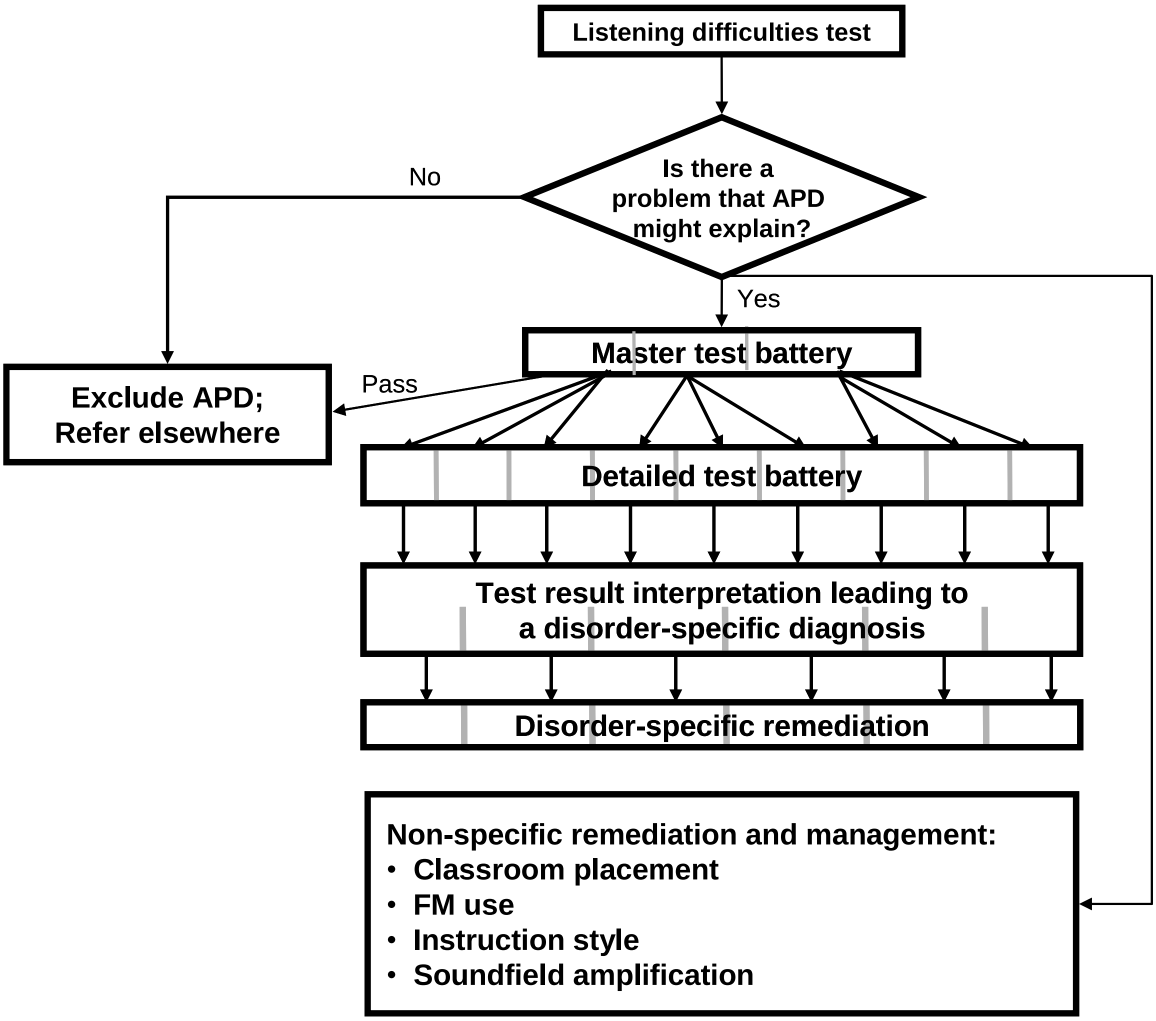


Figure 1. Model of a hierarchically structured adaptive, testing, remediation and management process for people reporting difficulties hearing in difficult listening conditions.

Following a questionnaire, or history taking, to confirm the presenting concerns, and an audiogram and immitance test, to determine the state of the peripheral hearing mechanism (outer, middle and inner ear), the first level of APD assessment comprises a test of listening difficulty. We note that an APD is possible irrespective of the outcomes of the peripheral hearing assessment. Indeed, recent results at our laboratory indicate that the presence of a sensorineural loss almost guarantees that spatial processing disorder (a specific form of APD) is present (Glyde, Cameron, Dillon and Hickson, in preparation).

A test of listening difficulties independently confirms the self-reported, or parent/teacher-reported, difficulties. This test has to have characteristics that enable the score to reflect the ease or difficulty the child has, relative to the average child of this age, in understanding speech in difficult but realistic listening conditions, irrespective of the cause of the disorder. Important factors to include in such a test would be realistically rapid speech, reverberation, competing signals, linguistically redundant cues, spatial cues, pitch cues, and demands on auditory working memory. As an alternative to a single test of listening difficulties, a small number of screening tests sensitive to different causes of hearing difficulty could be performed.

If the listening difficulty test confirms that a disability exists, then one is justified in performing a limited test battery without statistical correction for performing multiple tests. Interpretation without a correction for multiple tests can be justified because one has already established that there is a problem, so a fail result on one of a small number of tests is more likely to be because there is a problem than because of the accumulating probability of a fail when multiple tests are used. This is analogous to performing a planned-contrast statistical analysis once the main effect has been shown to be significant. Looked at another way, many false fails are prevented, because the battery is not administered to the many children for whom a listening difficulty cannot be demonstrated in the first level assessment.

The size of the master battery cannot, however, be increased without limit. As the size is increased, so too is the probability of a fail for reasons unrelated to the patient's real-life communication ability. It seems likely that the most effective rehabilitation training will be training that is focused on the specific disorder present for that patient. If specific disorders (of which there may be many) each can be detected only with a specific test or test condition, how can we perform a multitude of tests to determine the specific nature of the disorder for each patient, without suffering all the disadvantages of a large test battery?

One option would be to design a master test battery so that each test in the master battery produces a score below some cut-off value whenever any one of a limited number of disorders is present to a sufficient degree. An adverse result on such a test would then not only confirm that there is a problem impacting on listening, but also suggest which further, more detailed, tests should be administered. Each of those more detailed tests must then be focused on one of the disorders that might cause a low score in the corresponding test in the master test battery. The detailed tests could be speech-based, or psychoacoustic in nature, or a mixture of the two. Depending on which one of the tests in the master test battery gave a below-normal score, a different set of detailed tests would be administered. This in effect narrows the field of search to enable the clinician to efficiently identify the specific deficits contributing to the listening difficulty, without needing to employ the entire test battery.

Having determined the specific disorder or disorders present, disorder-specific remediation can then be given. The benefit of remediation can be assessed by repeating the same chain of tests: the disorder specific test, the individual test from the master battery that is sensitive to that disorder, and the high-level measurement of listening difficulty that better simulates real-life understanding of speech. Demonstration of benefit on the high level listening difficulties test is important. Disorder-specific remedial training is likely to involve tasks similar to those the child had difficulty with in the most detailed diagnostic tests (for example, practicing memorizing, recalling and manipulating items or sentences if a problem with auditory working memory was diagnosed). Improvement on the high-level listening difficulty test is needed to demonstrate, in the clinic, that the improved skills have generalized in a manner useful in real-world listening conditions. The time frame following remediation for a demonstrable improvement in a specific deficit versus a demonstrable improvement in overall listening difficulties may not necessarily be the same, however.

For any of the tests to be used to assess improvement following remediation, they must have strong test-retest reliability and documented one-sided critical difference scores. These scores are required to determine whether an individual has significantly improved on a particular diagnostic test following remediation or compensation, taking into account both measurement error, and the mean test-retest differences that occur in the absence of training. For example, for an improvement in an individual score to be significant at the 95 percent confidence interval level, the score must increase by 1.64 times the standard deviation of the test-retest difference plus the mean test-retest difference (Cameron and Dillon, 2007).

Note that non-specific management advice regarding classroom placement, teacher instruction style, and FM or soundfield usage can be given after the first level of assessment, as this advice relies only on a problem being present; knowledge of its underlying cause, including whether an APD is present, is not needed. These management strategies all attempt to decrease the difficulty of the listening condition, which is appropriate if the patient has more difficulty than the patient's peers. It is common practice to provide this management advice irrespective of the outcome of the APD test battery, though this lack of connection with the test result is not always acknowledged. The proposed model explicitly recognizes the lack of connection between generalized management and the APD test battery results.

Tests to fill all of the boxes in Figure 1 do not exist as yet, though many established tests could find a place within the structure. It is our goal to progressively create sufficient tests to enable such structured testing to take place. We publish this article in the hope that other researchers will also think about how tests they create might fit into a systematic structure that overcomes the problems of multiple tests but enables many specific causes of listening difficulties to be identified.

**Hierarchical interpretation example – the LiSN-S**

As an illustration of tests that would readily fit into the structure of Figure 1, we will show how the Listening in Spatialized Noise – Sentences test (LiSN-S), can be interpreted in a hierarchical manner. The LiSN-S consists of four sub-tests. These are designed to be administered in a standard order (to which the normative data applies), but interpretation is none-the-less hierarchical. Each sub-test comprises simple target sentences presented binaurally under headphones in the presence of distracting voices. The level of these target sentences is varied adaptively to find the signal-to-noise ratio (SNR) at which 50% of the words in the target sentences are understood. This SNR is referred to as the speech reception threshold in noise ($SRT_n$).

The stimuli, both the target sentences and the distracters (continuously presenting children's stories), have been processed with head-related transfer functions (HRTFs) to give them an apparent direction of arrival. Target voices always appear to arrive from the front. In two of the subtests, the distracters also arrive from the front, and in each of the other two subtests, one distracter appears to arrive from the right side of the head and the other distracter appears to arrive from the left side of the head. Pairs of subtests with the same distracter directions of arrival differ in talkers used. In one subtest the distracting stories were recorded by the same female talker as for the target sentences (shown shaded in Figure 2), whereas in the other subtest, they were recorded by two different female talkers. The four subtests are shown in Figure 2.

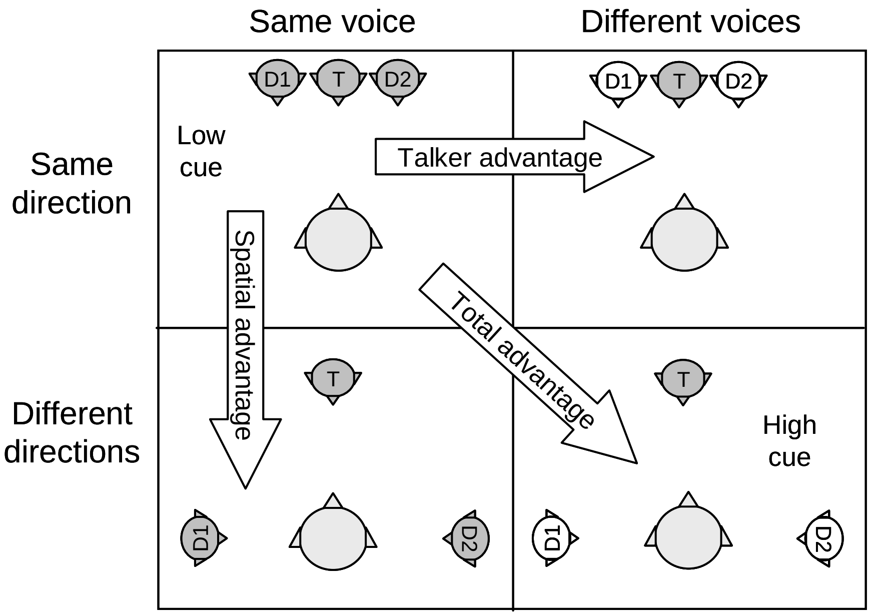


Figure 2. The four subtests of the LiSN-S test, and the three difference scores (advantage measures) that can be derived from them. The target speech, T, always comes from the front, whereas the two distracters, D1 and D2, come from the front or the sides, in different conditions.

From these four sub-tests, three difference scores can be calculated:

- Talker Advantage, which shows the amount in dB by which the SRT can be improved by using information about the different voice qualities (such as timbre and fundamental frequency contour) of the distracters relative to the target voice.
- Spatial Advantage, which shows the amount in dB by which the SRT can be improved by using information about the different spatial locations of the distracters relative to the target.

- Total Advantage, which shows the amount in dB by which the SRT can be improved by combining information about the different voice qualities and the different spatial locations.

Additionally, two of the individual sub-test scores are particularly valuable:

- The low cue SRT shows the SNR needed when there are no talker or spatial cues available.
- The high cue SRT shows the SNR needed when there are both talker and spatial cues available.

Of these five measures, the measure affected by the widest range of potential disorders (but certainly not by all potential disorders) is the high cue SRT. If the high cue score is sufficiently close to normal, we can rule out the child having a problem with using spatial cues or talker cues and not test the remaining three conditions. We can regard the high cue condition as one of the tests in the master battery and the remaining three conditions as part of the detailed battery, delivered only when performance is sufficiently poor on the high cue condition. Pass-fail criteria for each of the five scores have been defined in the usual way, as the value 2 S.D.s below the age-adjusted mean performance of children with no known listening difficulties, though we are the first to acknowledge the inadequacy of defining a pass-fail criterion in this way.

Over a number of studies, we have administered the LiSN-S to 183 children who have been identified by either teacher or parent report as having difficulty understanding speech in background noise. Of these, 40 children (22%) failed the high cue measure. The proportion, though much bigger than the 2.5% expected to be below a cut-off score that is 2 S.D.s below the mean for children with no reported listening difficulties, is relatively small as the test has been designed to be minimally affected by poor auditory memory, language knowledge, ability to understand instructions, and attention. None-the-less, poor performance on this measure ***could*** be a result of a grossly inadequate vocabulary, very poor short-term memory, poorly formed internal templates of how phonemes and/or words should sound, poor attention, inability to understand the task, or what we traditionally recognize as an APD, including a temporal processing disorder (such as reduced ability to perceive gaps or to perceive or recall the order of stimuli), reduced ability to use spatial information, reduced ability to use pitch differences, reduced ability to use timbre differences, and reduced ability to use redundant information (closure skills).

Many of these issues are not regarded as auditory processing skills as they involve abilities related to language, attention or cognitive function, and there may or may not be a correlated ability that can be assessed by a visual test. However, there is no need to go down the thorny, and in the end semantic, path of what we mean by auditory processing. Any of these reduced abilities, if severe enough, will make it more difficult for the child than for his or her peers to understand speech in difficult listening conditions.

Now consider what we can learn from the other subtests and difference scores when the high cue score is abnormal. If the low cue score is normal, most of the potential causes of the problem listed in the previous paragraph can be ruled out. Both sub-tests have the same requirements for almost all of the potential causes listed, so a problem with any of these is very unlikely to cause an aberrant score on the high cue subtest but not on the low cue subtest. By contrast the ***difference*** between the two scores, tells us specifically about the child’s ability to combine spatial and talker-related information. We can go further: the spatial advantage measure is the difference between the SRTs on two sub-tests that differ only in the spatial location of the talkers. Consequently, an aberrant spatial advantage score must indicate a deficiency in ability to use spatial cues to segregate a target talker from distracting talkers. Similarly, an aberrant talker advantage score must indicate a deficiency in using the differences in voice quality to segregate a target talker from distracting talkers.

The fact that it is possible to devise a wide range of tests in which each test, or combination of tests, assesses a specific disorder does not mean that all these tests will be equally useful. Of the 183 children with reported listening difficulties to whom we have administered LiSN-S, 32 have proven to have a spatial processing disorder. This disorder is characterized by a pattern of depressed scores (currently defined as more than 2 S.D.s below age-adjusted average) in those conditions where the target and distractors are spatially separated, but within ± 2 S.D.s of age-adjusted average when the target and distractors are spatially coincident. In contrast, only 4 children have proven to have a talker advantage score more than two S.D.s below normal. Consequently, it appears that spatial processing disorder is a significant cause of listening difficulty in the classroom, whereas talker discrimination (which is presumably enabled by perception of pitch and/or perception of timbre) is not a problem that many children have.

Note that although all of the LiSN-S subtests are language based, there is no chance that language based-disorders (i.e. a specific language impairment of some type) can cause any pattern of test results that would lead to a mistaken diagnosis of spatial processing disorder. The first defense against a language impairment causing poor performance was the choice of stimuli: They were specifically developed to be familiar to young children (from age 4) with a hearing loss. The second, and more important defense, which guards against the effects of attention, auditory memory, and many other potentially confounding factors, arises from the principle of using difference scores, derived from subtests that differ in only one respect. If two subtests differ in only one respect, then any difference in scores between those subtests can be the result only of factors affected by that specific difference in test conditions, plus ever-present measurement error.

In the case of the LiSN-S test, the same demands are placed on language ability, working memory and attention irrespective of whether the target and competing sounds are spatially co-located or spatially separated. Another example of differential analysis is the pitch pattern sequence test (Pinhiero, 1977) in which the same stimuli are heard in two subtests, but the responses required are imitation in one subtest and verbal labeling in the other subtest. A significant difference in scores can arise from the processes involved in imitation or those involved in labeling, but not from the many processes involved in detecting the pitch difference between the signals which are common to the two subtests. Similarly, Lagace, Jutras and Gagne (2010) suggest that use of the Speech Perception in Noise (SPIN) test (Kalikow et al., 1977) with its high and low context sentences, enables the cause of increased difficulty to be split between factors related to language and factors related to auditory abilities. This differential approach could be a good basis for developing further tests that can identify specific auditory abilities for which performance is outside the normal range.

Test reliability is of critical importance when difference scores are used. When two test scores (e.g. SRTs) are subtracted, the measurement error associated with each of the subtest scores are added (or more correctly, the error variances are added). Consequently, each subtest must have a small random error component or else the difference score will have an error so large that the difference score lacks sensitivity to any disorder. Measurement error can be kept small by a combination of precise calibration, appropriately equalized test stimuli, test technique and test brevity that maintains a child's attention, and adaptive test procedures that cause each trial to provide useful information, rather than some of the trials being wasted on conditions that give floor or ceiling performance. Continuous calculation of test precision as each subtest progresses also helps as the subtest can be truncated when the requisite precision has been reached.

**Overall focus**

We believe that the overall goal of clinicians should be on determining whether a patient has a problem listening in difficult listening conditions and if so, discovering in as much detail as possible the reason for the problem. If a specific cause can be found, there is the possibility of providing disorder-specific remediation. Providing the tests that enable clinicians to do this seems an appropriate goal for researchers.

This focus on determining why speech perception is poor in difficult listening conditions can apply whether the cause lies in the realm of auditory processing, or some other skill. In principle, the master tests that sit on the second level in Figure 1 could include tests sensitive to factors like language, attention, and memory, factors that would generally be considered to lie outside of auditory processing but which may none-the-less cause poor speech perception in adverse listening conditions.

Such a focus on solving the problem of difficulty in understanding speech seems more productive than trying to define APD, or equally to classify a particular patient as "having", or "not having", APD. Many unproductive arguments can be avoided:

- whether the name should be APD, CAPD or (C)APD;

- whether difficulties such as poor auditory memory, or poor stored representations of phonemes, should be classified as APD or as something else;
- whether a disorder can be called APD if the patient also has below-normal performance on a seemingly similar task in another sensory modality;
- Whether a disorder can be called APD only if it has only been diagnosed with a non-speech test, a speech test, or both;
- whether a specific site of lesion has yet been discovered where processing underlying the specific ability normally takes place.

None of this problem-solving approach for clinical practice takes away from the need to better understand the complete auditory system – how it normally works, what can go wrong, where this is happening within the auditory system, and what the consequences of these disorders are. Development of models of APD that both recognize the heterogeneous nature of APD and describe the disorder in terms of patterns of findings that frequently co-occur will be important in advancing our knowledge of auditory processing mechanisms. Some models already exist (e.g. Bellis and Ferre, 1999; Katz, 1992) but have been criticized as lacking adequate data and producing inconsistent results as to which children "have" APD, and what the nature of that disorder is (Jutras et al., 2007).

We commend the development of further tests and remediation strategies that will facilitate the identification and remediation of the various causes of difficulty understanding speech in noise. A worthy first goal would be the creation of a high-level test of listening difficulty that simulated listening in difficult listening conditions and in which performance was measurably affected for anyone who has impaired ability to understand speech in difficult listening conditions in real life, irrespective of the cause. Some of the characteristics of such a test might include:

- High rate of presentation, and some reverberation – to create difficulties for people with reduced temporal processing ability;
- Spatial separation of target and competition – to create difficulties for people with reduced spatial selectivity;
- High-context material – to create difficulties for people with reduced auditory closure skills;
- High memory load – to create difficulties for people with reduced auditory working memory;
- Contrasting pitch contours in the competing sounds – to create difficulties for people with reduced pitch perception.

Such a test that was affected by a wide range of abilities would in no way be diagnostic of the cause of the problem, but would be very valuable in confirming that there was a problem, and in measuring its magnitude.

Finally, while we have not mentioned electrophysiological tests in this article, we expect that they will have a growing role in assessment of children who report listening difficulties, but they are outside the scope of this article.

## Acknowledgments

We would like to thank Suzanne Purdy and Mridula Sharma for their helpful comments on an earlier version of this manuscript.

## References

American Academy of Audiology (2010). American Academy of Audiology Clinical practice guidelines: Diagnosis, treatment and management of children and adults with central auditory processing disorder. Available from http://www.audiology.org/resources/documentlibrary/Documents/APD Guidelines 8-2010.pdf

American Speech-Language-Hearing Association (2005). (Central) auditory processing disorders [Technical Report]. Available from http://www.asha.org/docs/html/TR2005-00043.html#sec1.11.

Bamiou DE, Free SL, Sisodiya SM, Chong WK, Musiek F, Williamson KA, van Heyningen V, Moore AT, Gadian D, Luxon LM. (2007) Auditory interhemispheric transfer deficits, hearing difficulties, and brain magnetic resonance imaging abnormalities in children with congenital aniridia due to PAX6 mutations. *Arch Pediatr Adolesc Med* 161(5):463-9.

Baran JA. (2007) Test battery considerations. In F.E. Musiek and G.D. Chermak (Eds.), *Handbook of (Central) Auditory Processing Disorder: Auditory Neuroscience and Diagnosis* (Vol 1, 163-192). San Diego: Plural Publishing.

Bellis TJ, Ferre JM. (1999) Multidimensional approach to the differential diagnosis of central auditory processing disorders in children. *J Am Acad Audiol*, *10*(6):319-328.

Boscario M, Garcia VL, Guimaraes CA, Montenegro MA, Hage SR, Cendes F and Guerreiro MM. (2010). Auditory processing disorder in perisylvian syndrome. *Brain Dev*. 32(4):299-304.

Cameron S, Dillon H. (2007) The Listening in Spatialized Noise - Sentences Test (LISN-S): Test-retest reliability study. *Int J Audiol, 46:* 145-153.

Cameron S, Dillon H. (2008) The Listening in Spatialized Noise - Sentences Test: Comparison to prototype LISN test and results from children with either a suspected (central) auditory processing disorder of a confirmed language disorder. *J Am Acad Audiol*, 19(5): 377-391.

Cameron S, Dillon H. (2009) *Listening in Spatialized Noise – Sentences test (LiSN-S)* (Version 1.014) [Computer software]. Murten, Switzerland: Phonak Communications AG.

Glyde H, Cameron S, Dillon H, Hickson L (In preparation) The effects of hearing impairment and ageing on spatial processing.

Hurley RM, Musiek FE (1997). Effectiveness of three central auditory processing (CAP) tests in identifying cerebral lesions. *J Am Acad Audiol* 8(4):257-62.

Jutras B, Loubert M, Dupuis JL, Marcoux C, Dumont V, Baril M. (2007) Applicability of central auditory processing models. *Am J Audiol* 16:100-106.

Kalikow DN, Stevens KN, Elliott LL. (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *J Acoust Soc Am* 61(5):1337-51.

Katz J. (1992) Classification of auditory processing disorders. In Katz J, Stecker NA, Henderson D (Eds.), *Central auditory processing. A transdisciplinary view (pp. 81-91).* St Louis: Mosby Year Book.

Lagace J, Jutras B, Gagne JP. (2010) Auditory processing disorder and speech perception problems in noise: finding the underlying origin. *Am J Audiol* 19(1):17-25.

Musiek FE, Bellis TJ, Chermak GD (2005). Nonmodularity of the central auditory nervous system: implications for (central) auditory processing disorder. *Am J Audiol* 14(2):128-38.

Musiek FE, Gollegly KM, Ross MK (1985). Profiles of central auditory processing disorders in children with learning difficulties. *Commun Disord Quart* 9(1):43-63.

Musiek FE, Pinheiro ML, Wilson DH (1980). Auditory pattern perception in 'split brain' patients. *Arch Otolaryngol* 106(10):610-2.

Musiek FE, Sachs EJr (1980). Reversible neuroaudiologic findings in a case of right frontal lobe abscess with recovery. *Arch Otolaryngol* 106(5):280-283.

Musiek FE, Kibbe K, and Baran JA (1984). Neuroaudiological results from split-brain patients. *Seminars in Hearing* 5:219-229.

Pinhiero M L (1977). Tests of central auditory function in children with learning disabilities. In R. W. Keith (Ed.), *Central auditory dysfunction* (pp. 223-256). New York, NY: Grune & Stratton.

**Figure Legends**

Figure 1. Model of a hierarchically structured adaptive, testing, remediation and management process for people reporting difficulties hearing in difficult listening conditions.

Figure 2. The four subtests of the LiSN-S test, and the three difference scores (advantage measures) that can be derived from them. The target speech, T, always comes from the front, whereas the two distracters, D1 and D2, come from the front or the sides, in different conditions.