2010-06-29

Normal Masking Level Difference Parameters For Use in the Clinical Evaluation of Auditory Processing Disorders

Maria Noelle Burnham

Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Communication Sciences and Disorders Commons

BYU ScholarsArchive Citation

Burnham, Maria Noelle, "Normal Masking Level Difference Parameters For Use in the Clinical Evaluation of Auditory Processing Disorders" (2010). All Theses and Dissertations. 2139.
https://scholarsarchive.byu.edu/etd/2139

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen amatangelo@byu.edu.
ABSTRACT

Normal Masking Level Difference Parameters For Use in the Clinical Evaluation of Auditory Processing Disorders

Maria N. Burnham
Department of Communication Disorders
Master of Science

Masking Level Difference (MLD) tests are an established component of auditory processing test batteries; however, normative data for these tests vary according to procedure. In this study, forty normal, native-English speaking adults between the ages of 18 and 26 were tested for MLD via a newly developed computer software program using both an adaptive procedure (MLDA) and a Bekesy procedure (MLDB). The results from the two procedures were analyzed for sex differences and compared with each other.

For both the MLDA and MLDB, the results showed statistically significant sex differences in the masked thresholds used to obtain the MLD (N_o S_o and N_o S_π), but no significant difference in the calculated MLD value (N_o S_o - N_o S_π). These results suggest that since the MLD was similar for both sexes, the normative data need not be reported separately by sex.

The results also showed statistically significant differences between procedures, with the MLDA procedure producing higher MLDs than the MLDB procedure. The MLDA procedure lent itself to a d’ analysis, which could not be determined using MLDB due to the nature of a Bekesy assessment. For MLDA, d’ = 1.4, test sensitivity = 96.4%, and test specificity = 60.3%. The results of this study indicate that MLDA is a better testing procedure due to MLDA’s higher MLD average and the statistical data available (d’, and measures of sensitivity and specificity) when using the MLDA procedure.

Keywords: Masking level difference, MLD, adaptive, auditory processing disorders, masked thresholds, Bekesy
ACKNOWLEDGMENTS

I would like to thank my parents for their continuing support and encouragement, and all my family members who have housed me, fed me, and listened to my success and frustrations over the past three years. Aunt Ardie and Uncle Heber, and Lincoln and Donna have earned special thanks for their exceptional kindness this past year.

I would also like to thank Amanda Fujiki for her help with the data collection and Andrzej Senderski for his help in setting up the testing software and running the preliminary tests. Completing this project would have been impossible without the guidance and support of Dr. McPherson, my thesis chair. Thank you so much for you time, expertise, and kindness. It has been a pleasure working with you.
### Table of Contents

Introduction ........................................................................................................................................... 1

Review of Literature ................................................................................................................................. 5

Current Models for APD Diagnosis .......................................................................................................... 5

- The Buffalo Model ................................................................................................................................. 5

- The Three-Test Battery .......................................................................................................................... 5

- Four Categories of APD .......................................................................................................................... 7

- The Bellis/Ferre Model .......................................................................................................................... 8

- Criticism of the Buffalo and Bellis/Ferre Models .................................................................................. 11

Masking Level Difference Tests ............................................................................................................. 12

Method ....................................................................................................................................................... 17

- Present Study ......................................................................................................................................... 17

Participants .............................................................................................................................................. 18

Apparatus ................................................................................................................................................. 19

Procedures ............................................................................................................................................... 19

- Adaptive Masking Level Difference ..................................................................................................... 20

- Bekesy Masking Level Difference ......................................................................................................... 22

Results ......................................................................................................................................................... 23

- Descriptive Data ...................................................................................................................................... 23

- Inferential Data ....................................................................................................................................... 23
List of Tables

Table                                                                 Page

1. Average thresholds and MLDs for Adaptive (MLDA) and Bekesy (MLDB) procedures (female = 20, male = 20, combined = 40) ................................................................. 24

2. Average d’ for Adaptive Masking Level Thresholds ................................................................. 26
<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Receiver Operating Characteristic (ROC) curve for the combined female and male participants for the MLDA Test.</td>
<td>27</td>
</tr>
</tbody>
</table>
## List of Appendixes

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Interview Questions</td>
<td>40</td>
</tr>
<tr>
<td>B. Informed Consent</td>
<td>41</td>
</tr>
</tbody>
</table>
**Introduction**

Processing disorders of the central auditory system have been recognized since the mid-twentieth century (Lucker, 2007). Unlike the more common peripheral auditory disorders in sound perception, auditory processing disorders (APD) cannot be diagnosed with pure-tone audiometry, routine speech audiometry, or other standard audiological tests of threshold function. This is because the breakdown in auditory processing occurs in the central nervous system and not in the conductive or sensory auditory system. Individuals with APD are able to hear, but find it difficult to process language into meaningful information or understand the sounds they are hearing (Berry & Eisenson, 1956; Lucker, 2007). The most significant result of this disorder is impaired communicative ability. The extent of the communication impairment, as explained by Lucker (2007), is a spectrum that can range from receptive language difficulty to a complete inability to acquire language.

The relationship between auditory processing and higher cognitive function is illustrated by the observation that APD has been associated with speech disorders, learning disorders, and social problems (Bellis 2002; Bellis & Ferre, 1999; Chermack & Musiek, 1997; Katz, 1992). Although for individuals with mild impairment, APD may not necessarily pose a significant problem; yet, for individuals with a more significant impairment, the quality of life may be greatly impeded. Consequently, accurate assessment and identification of APD is essential for accurate diagnosis and treatment of APD. Finding tests and models of APD that are both reliable and efficient has proved to be a difficult task.

One reason APD diagnosis is difficult is that, by definition, auditory processing includes a variety of independent and integrated skills. Auditory processing skills include sound localization and lateralization, auditory discrimination, pattern recognition, temporal integration,
temporal discrimination, temporal ordering, temporal masking, listening comprehension with competing noise, and listening comprehension with degraded acoustic signals (American Speech-Language-Hearing Association [ASHA], 2005; Bellis, 2003; Geffner; 2007). Difficulties with any of these skills can result in a diagnosis of APD. Thus, tests of auditory processing function need to probe several different processing skills.

Diagnosing APD is complicated by the fact that APD often co-occurs with other disorders. Attention deficit hyperactivity disorder (ADHD) is one of the most common disorders to occur with APD (Geffner, 2007; Hamaguchi & Tazeau, 2007). Differentiating between APD and ADHD is difficult because they not only co-occur, but also share many of the same identifying behaviors (Bellis, 2003, 2007; Chermak, 2007; Geffner, 2007; Hurley & Hurley, 2007). For example, distractibility, avoiding difficult undertakings, and appearing not to listen when spoken to are all hallmarks of both ADHD and APD. Many of the same identifying behaviors are also characteristic of auditory neuropathy (Starr et al., 1991), non-auditory causes (Geffner, 2007) and peripheral hearing loss (Lucker, 2007). In addition, each of those disorders is known to frequently be associated with learning, speech, and social difficulties (Bellis, 2003, Geffner, 2007). Two or more of these conditions often occur together; however, they may also all occur independently. Thus, the presence of one disorder cannot be used to diagnostically confirm or rule out the presence of another disorder. Differential diagnosis of APD amid associated problems makes it difficult to identify individuals or populations with APD. Since there is a lack of a valid standard for diagnosing APD, this results in confusion in its diagnosis.

In order to identify APD separately from other disorders it is necessary to use tests that target the central auditory processing system as directly as possible. There are a number of psychoacoustic skills that are representative of the auditory processing system. For each skill
there are several auditory tests that can access and assess the specific psychoacoustic functions that represent auditory processing abilities. However, when using these tests to diagnose APD, there is no gold standard available to identify individuals with APD. A gold standard is a way to compute test efficiency by collecting measureable data from both those who are free of the disorder as well as those who already have a documented diagnosis, thus establishing normative data for both conditions. Unfortunately, “because of the variability and the nature of the profiles of (C)APD, there exists no absolute gold standard for deriving sensitivity and specificity data for tests of central auditory dysfunction” (ASHA, 2005, p. 11).

The problem of identifying a population with APD in order to establish a gold standard refers to the earlier discussion of co-occurring and associated problems. Establishing an APD population based on behavioral symptoms would be problematic since APD behaviors are also associated with many other common disorders. The same problem occurs when trying to establish an APD population using children with learning disabilities (Bellis, 2003). Children, however, are unable to perform complicated psychoacoustic tests, which further aggravates the issue. As a result of the problem with establishing a gold standard, most tests of central auditory function lack significant normative and standardized data (Cacace & McFarland, 2005; Jutras, et al., 2007; Keith, 2009).

To be efficient, APD models should include multiple tests of auditory process that are based on psychophysical principles. In addition, these tests should be both user friendly and commercially available. The question of which tests should be included in an assessment of auditory processing ability has many differed answers, depending on the researcher. Some authors claim that only a few tests targeting the auditory processing system are enough to assess APDs (Jerger & Musiek, 2000; Katz, 1992). Other research indicates that, in order to fully
assess the complex auditory system, test batteries must include more tests to fully cover the potential deficits in the system (Bellis, 2003; Chermak & Musiek, 1997). The APD technical report published by ASHA (2005) clearly states that tests of multiple functions are needed, but does not specify a minimum number of tests or processes that should be covered, which leaves this issue somewhat open to debate.

There are many tests to choose from to create an APD test battery. It is common to include tests that use linguistic stimuli such as the Speech in Noise test, Stagger Spondaic Word test, or Dichotic Speech tests (ASHA, 2005; Bellis, 2005; Katz, 1992; Stollman et al., 2004). However, there are some who are concerned that the results of auditory processing can be confounded by the language-processing that occurs when decoding linguistic stimuli (Cacace & McFarland, 2005; Hurley & Fulton, 2007). Thus, it is also important to include non-linguistic auditory tests in any APD battery.

The need for more effective APD tests is recognized by makers of current APD diagnostic models (Bellis, 2003) as well their reviewers (Jutras et al., 2005; McFarland & Cacace, 2007). It is important to establish clear normative data and to cover a wide range of auditory processes. The creation of an easily administered APD test battery with clear normative data would greatly increase the ability to accurately diagnose APD in a time- and cost-effective manner. In the present study, a single test of central auditory processing, the Masking Level Difference (MLD), has been selected to evaluate an adaptive mode of testing suitable for further use in investigating APDs.
Review of Literature

Current Models for APD Diagnosis

Although there is no single universally accepted method of APD diagnosis, the two diagnostic models most commonly used are the Bellis/Ferre model and the Buffalo model. According to Bellis, the Buffalo model primarily focuses on how a child’s test performance compares to his or her specific learning difficulties; while the Bellis/Ferre model relies both on the relationship between test performance and learning difficulties within the context of the child’s underlying neurophysiology (Bellis, 2002; Bellis & Ferre 1999). In a review of the two different models, Jutras et al. (2007) indicate that both models refer to underlying neuroanatomy to explain APD.

The Buffalo Model. Katz (2007) at the State University of New York at Buffalo began developing what is now known as the Buffalo Model for APD classification in 1986. A description and explanation of the model was published in 1992 (Katz, 1992) and is one of the most widely used models for APD diagnosis. As described by Katz, the Buffalo model uses a three-test battery in conjunction with two surveys (one for the parents, one for the school teacher) in order to identify the areas of deficit in children with APD. The three tests used in the buffalo model are (a) Phonemic Synthesis, (b) Speech in Noise, and (c) the Staggered Spondaic Word (SSW) tests. The Buffalo Model explains not only what the test consists of, but also the types of processing skill dysfunctions that are associated with a low score.

The three-test battery. The Phonemic Synthesis test uses recordings of monosyllabic words presented at 50 dB SL. The words are broken into individual phonemes with a full 1.5 second pause between each phoneme (Katz, 2007). The patient must fuse these individual phonemes together in order to repeat the complete word. According to Katz (1992), the
Phonemic Synthesis test requires three different skills of the patient: the patient must be able to
(a) discriminate speech sounds, (b) remember which sounds were heard, and (c) integrate the
sounds into a whole word. Katz suggests that errors in Phonemic Synthesis can indicate
difficulties with decoding and memory.

The Speech in Noise test requires the patient to understand a speech signal presented at
the same time as a noise signal. This is a monaural test where speech is presented at 40 dB SL
and speech spectrum noise is presented at 30 dB SL in the monotic condition. According to Katz
(2007), the Speech in Noise test is clinically important since difficulty hearing in noise is a
common symptom of APD and is a function subtended by the inferior colliculus as well as
higher brain function.

In the SSW, spondee words are presented in the dichotic condition to the right and left
ears in a partially overlapping fashion. For example, the spondee green house is presented to the
right ear while string bean is presented to the left ear. If the right ear is leading, the patient will
hear green by itself in the right ear, followed by house and string heard simultaneously in right
and left ears respectively, and ending with bean heard by itself in the left ear. In the SSW, all
odd numbered items lead with the right ear, and all even numbered items lead with the left ear.
The patient’s answers are organized in such a way that for incorrect items, examiners can
identify not only the leading ear but also if the error-word was heard alone or in competition.
Thus, the results can be used to compare left ear to right ear, words heard alone to words in
competition, words heard alone at the beginning of the item to words heard alone at the end of
the item, and performance at the beginning of the test to performance at the end of the test.

Katz (1992) classifies three main types of errors as (a) Order Effects, (b) Type A
Patterns, and (c) Reversals. Order Effects occur when the patient either does better on the first
half of the SSW versus the second half of the SSW (low/high order effect) or does better on the second half of the SSW versus the first half of the SSW (high/low order effect). Katz’ 1978 manual on the SSW found that low/high effects were observed in patients with posterior temporal lobe lesions. Katz says that, due to the research associating the posterior temporal lobe with phonemic decoding and receptive language functions (Luria, 1966; Penfield, 1959), low/high order effects indicate poor phonemic decoding. In contrast, high/low order effects are attributed to damage in the frontal and anterior temporal lobes (as cited in Katz, 1992) and are associated with memory dysfunction. In addition, Katz (1992) suggests that the anterior temporal lobe and the ability to hear sounds in noise (Efron, Crandall, Koss, Divenyi, & Yung 1983) are associated with both high/low and low/high order effects, indicating problems hearing speech in noise as well as auditory memory deficits.

A Type A Pattern described by Katz (1992) occurs when a patient scores much lower on words in competition than on words alone. This pattern is associated with classic dyslexia. The severe reading and spelling disabilities of dyslexic patients is associated with disorders of the angular gyrus and posterior corpus callosum (Damasio & Damasio, 1983). Thus, a Type A Pattern can be an indication of reading and writing disorders possibly caused by damage to the angular gyrus or posterior corpus callosum.

Reversal Errors are when the patient reports the correct components of the stimulus, but in the wrong order. This can occur on the SSW, Phonemic Synthesis test or Sound in Noise test. Reversal Errors are associated with organization problems and damage to the mid-cerebrum (Luria, 1970). Errors of this type imply a specific sequencing disorder.

**Four categories of APD.** The three tests in the Buffalo model all provide important information for the classification of APD, however the foundation of the Buffalo Model is the
SSW test. From the information gathered with the SSW, Katz developed four separate categories of APD. Errors on the Phonemic Synthesis and Sound in Noise tests also indicate specific APD categories, but do not indicate additional categories from what was already established using the SSW. Katz’s four APD categories are (a) Decoding, (b) Tolerance-Fading Memory (TFM), (c) Integration, and (d) Organization.

According to Katz (2007), Decoding is “the ability to quickly and accurately digest speech, most importantly at the phonemic level,” (Decoding Category section, para. 1). High error peaks, low/high order affects in the SSW, and errors on the Phonemic Synthesis test are all indicators for the Decoding category. The TFM category is actually two categories in one: (a) the ability to understand speech in noise, and (b) short-term memory. High/low order effects and significant Speech in Noise scores are indicators for the TFM category. Integration is also broken into two parts: Type 1 and Type 2. Both types refer to disorders of reading and writing (auditory-visual integration) but Type 1 is more severe than Type 2 (Katz, 1992). Integration is indicated by Type A Pattern findings (Katz, 2007). Organization, the final APD category, reflects the patient’s ability to sequence accurately as well as direction following abilities (Katz 1992, 2007). Reversals on the SSW and the Phonemic Synthesis test are indicators for problems in the Organization category.

The Bellis/Ferre Model. Compared to the Buffalo Model, the Bellis/Ferre (Bellis, 2003) model for APD is much less structured. The differences between the two models can largely be explained by tenets of Bellis’ beliefs about APD testing. These tenets are (a) Test batteries should cover eight categories, (b) Each patient’s battery should be uniquely developed to highlight their particular strengths and weaknesses, and (c) APD testing is a dynamic process that requires clinicians to make test-choice decisions during the testing process in response to the
patient’s performance. Because of Bellis’ approach to testing, the Bellis/Ferre model neither chooses which tests to include in the battery nor gives a clear recipe for interpreting results. This model puts more responsibility on the clinician to be educated about auditory processes, available tests, and interpretations.

The eight categories of testing identified by Bellis (2003) as the components of a comprehensive APD test are (a) dichotic listening with directed attention, (b) dichotic listening reporting on both ears, (c) temporal patterning, (d) monaural low-redundancy speech, (e) temporal gap detection, (f) binaural interaction, (g) auditory discrimination, and (h) physiologic measures of auditory function. Bellis further states that one of the dichotic listening tests should be linguistically loaded, while the other test should carry a light linguistic load. Although a specific battery of tests is not elucidated, each category listed is an important component of the APD assessment and provides important information on the patient’s strengths and weaknesses. The patient’s test results are used not only to determine the presence or absence of APD, but also to develop a complete profile of the patient’s auditory processing strengths and weaknesses to help direct treatment and intervention.

The Bellis/Ferre model (Bellis, 2003) divides APDs into different categories. The Bellis/Ferre primary sub-profiles are (a) Auditory Decoding Deficit, (b) Prosody Deficit, and (c) Integration Deficit. The secondary sub-profiles are (a) Associative Deficit and (b) Output Organization Deficit. Auditory Decoding Deficit is associated with dysfunction of the primary auditory cortex in the language-dominant hemisphere. According to Phillips and Hall (1990), the neurons in the primary auditory cortex discharge with the onset of a stimulus with nearly as much precision as the auditory nerve and are important for processing the phonetic timing cues of speech such as voice onset time, spectro-temporal transitions, and place of articulation (Bellis
Damage to the primary auditory cortex most often affects the processes of auditory closure, temporal processing, speech-sound discrimination, and binaural separation/integration.

According to Bellis (2003), *Prosody Deficit* is associated with dysfunction of the right hemisphere and affects the processes of temporal patterning, auditory discrimination of nonspeech stimuli, and binaural separation/integration. *Integration Deficit* is associated with dysfunction of the corpus callosum. The posterior portion of the corpus callosum has been shown to be critical to the transfer of sensory information between hemispheres (Musiek, Reeves, & Baran, 1985). It has also been suggested that the corpus callosum plays an inhibitory role, which leads to lateralization and hemispheric dominance for certain tasks (Ringo, Doty, Demeter, & Simard 1994). The inhibition of neural responses due to competing ipsilateral stimuli (monotonic) could explain why damage to the corpus callosum results in great difficulty hearing the left ear during dichotic tasks (Milner, Taylor, & Sperry, 1968). Bellis (2003) reported that damage to the corpus callosum affects temporal patterning, binaural separation, and binaural integration abilities.

*Associative Deficit* is an inability to apply the rules of language to acoustic information and is associated with dysfunction in the left-hemisphere auditory areas. Research findings have shown the cerebrum to be left-hemisphere dominant for language function (Gazzaniga, 2002; Rasmussen & Milner, 1997), which explains why damage to this area often manifests as a language disorder. *Output-Organization Deficit* is the final subprofile. The area of cortical dysfunction is not yet certain but it results in difficulty planning, sequencing, or otherwise organizing responses to auditory information.

The Bellis/Ferre model (2003) uses knowledge of the CNS structures and function to make appropriate test choices for a test battery. The clinician must make educated decisions
about appropriate cut-off criteria for normal and non-normal test results and then must synthesize the test data with the information gathered regarding the patient’s academic, social, and language functioning. Although the Bellis/Ferre model provides many resources to help in these processes, much of the decision-making process is left up to the individual clinician. With a comprehensive test battery, clinicians using the Bellis/Ferre model should be able to gather enough data to create a complete profile of the patient’s strengths and weaknesses. This information gives a holistic picture of the patient’s needs and can be extremely valuable when planning intervention strategies.

**Criticism of the Buffalo and Bellis/Ferre Models.** According to Jutras et al. (2007), the main difference between the two models is that the Buffalo model relies heavily on performance on the SSW test where the Bellis/Ferre model incorporates several different tasks including the SSW and various listening tasks in noisy conditions.

In a retrospective study, Jutras et al. (2007) assessed how well the two models actually identified APD populations. For their assessment, Jutras et al. used 178 children who had been diagnosed and were receiving treatment for APD. Accuracy of the initial diagnosis was difficult to confirm, but Jutras et al. reported that at least 60% of the children met the ASHA disorder criteria of at least two tests with scores more than 2 standard deviations below the mean (ASHA, 2005). Using the test scores collected in the initial diagnosis, Jutras et al. re-evaluated each child’s information using both the Bellis/Ferre model and the Buffalo model. The results showed that the Buffalo model identified 87% of the children with APD while the Bellis/Ferre model identified only 6%. These results indicate that while the Bellis/Ferre model may be useful for identifying specific strengths and weaknesses in children with APD, it is not clinically useful for diagnosis. It is also important to note that, while the Buffalo model achieved a much higher
percentage of diagnosis, there are still concerns regarding this model since it relies so heavily on one test which is heavily linguistically loaded. Without including more tests to describe the range of skills affected by APD, using the Buffalo model alone would not assess all of the auditory processing functions associated with APD (Bellis & Ferre, 1999; Jutras et al., 2007). In their conclusion, Jutras et al. stated that both the Bellis/Ferre and Buffalo models were “inadequate for clinical use and must be further refined” (2007, p. 105).

**Masking Level Difference Tests**

Individuals with APD often complain of difficulty hearing in noise. This difficulty is also connected with several other common symptoms associated with APD. In a list of classroom behaviors that are considered red flags of APD in children (Bellis, 2003; Geffner, 2007), nearly half of the behaviors could be directly related to the ability to hear in a noisy environment. The behaviors that are associated with hearing in noise identify children who

• can learn through the auditory channel but do better with visual stimuli,
• cannot write from dictation,
• mishear words,
• don’t participate in class discussions,
• have problems learning a foreign language,
• misunderstand homework assignments,
• fail to follow directions,
• cannot tolerate a noisy room,
• are fidgety in loud/noisy places such as the cafeteria, gym, or playground, and
• appear to have a delayed response to questions.
A test that directly assesses a patient’s ability to hear in noise is the masking level difference (MLD) test. Masking is often used during audiological testing to evaluate the effects of noise on auditory threshold and function. This is done by simultaneously presenting a masker (noise) with a tone or speech-sound (signal) and then adjusting the intensity of the signal to find the point at which the patient can hear the signal 50% of the time (threshold). In early masking tests, Hirsch (1948), Licklider (1948) and Webster (1951) observed that masking thresholds varied significantly with changes in phase relationship between noise and signal.

Different phase conditions are a result of adjusting the phase relationship between noise and signal; the noise may be presented either binaurally in-phase, or binaurally 180° out-of-phase and the signal may be presented monaurally, binaurally with both ears in-phase, or binaurally with the ears 180° out-of-phase. When the noise and signal are in-phase the condition is said to be homophasic. Research has repeatedly shown that masking is more effective in the homophasic condition (Chermak & Musiek, 1997; Harris, Brey, Miller, & Channell, 1992; Licklider, 1948; Olsen, Noffsinger, & Carhart, 1976). Abbreviations for each possible condition of noise and signal have been developed and are widely used. Abbreviations for the different binaural homophasic conditions include noise and signal in-phase (NₒSₒ), and noise and signal out-of-phase (NₒSₚ).

When the noise and signal are out-of-phase, the stimulus is said to be antiphasic, and as such, the masking is much less effective. This makes the patient’s threshold decrease significantly. Abbreviations for the different binaural antiphasic conditions include noise in-phase with signal 180° out-of-phase (NₒSₚ), and noise 180° out-of-phase with signal in-phase (NₚSₒ).
The change from effective masking in a homophasic condition to the less effective masking in the antiphasic condition is called release from masking (Olson 1976). The amount of masking release that occurs is measured by subtracting one masked threshold from another. This difference is called the MLD. Although many studies collect and report differences for multiple threshold combinations, (e.g., $N_oS_o - N_oS_\pi; N_oS_\pi - N_\pi S_o; N_oS_o - N_\pi S_\pi$), the MLD is calculated by subtracting $N_oS_\pi$ from $N_oS_o$. The $N_oS_\pi$ condition provides the greatest release from masking while the $N_oS_o$ condition provides the least release from masking (Hirsh, 1948). Thus, the difference between these two conditions ($N_oS_o - N_oS_\pi$) results in the largest possible MLD (Olsen et al., 1976) and is the value referred to as MLD in the current study.

The MLD has been reported with a range from 8.2 dB to 13.7 dB (Harris et al., 1992; Jerger et al., 1984; Olsen et al., 1976; Quaranta & Cervellera, 1977; Wilson, Arcos, Brown, & Bennett, 1984). The variation in MLD occurs due to differences in testing procedure such as (a) differences in noise and signal combinations, (b) differences in type of masking noise, and (c) length of signal. Examiners have several options when choosing the type of noise and signal combination. MLD examiners may use a sinusoid noise with a sinusoid signal, a broad or narrow-band noise with a speech signal, or a broad or narrow-band noise with a sinusoid signal. MLDs are fundamental to the assessment of APD since the ability to understand speech in noise is functionally related to everyday communication problems often experienced by those with APD. However, in a study comparing tonal MLDs with speech MLDs, Sweetow and Reddell (1978) reported no significant difference in the speech MLD between normal children and children with suspected APD. By contrast, tonal MLDs separated normal children from children with suspected APD with 79% sensitivity and 88% sensitivity. These results indicate that
although speech MLDs may provide important descriptive information on the patient’s functional abilities, tonal MLDs are likely to be more valuable for APD diagnostic purposes.

Masked threshold normative data have been reported for both amplitude-modulated noise and filtered-random noise. Harris et al. (1992) found that MLDs were larger by 2.4 dB for amplitude-modulated noise than for filtered-random noise, which was found to be significant; $F(1,99) = 46.8, p < 0.0001$. It was also found that larger MLDs were obtained when a continuous masking noise was used, as compared to noise with a center frequency equal to that of the sinusoid signal (Green 1965), and that larger antiphase thresholds occur when the stimulus intensity is increased (Townsend & Goldstein, 1972). Similarly, Green (1965) found that signal durations of 10 ms resulted in MLDs approximately 2 dB larger than for signal durations of 100 ms and 1000 ms. These variations in MLD testing account for the variations in MLD values obtained by different experiments. The implication of these variations is that clinicians must be careful to compare test results to normative data that have been collected using similar methods.

The MLD is a binaural interaction phenomenon requiring the CNS to process and combine signals from both ears (Bellis, 2003). The MLD test examines interaural differences in auditory processing of intensity and timing of signals received simultaneously by the two ears (Johnson, Bellis, & Billiet, 2007) and reflects localization ability (Gelfand, 2001; Webster, 1951). Hirsch (1948) first described the effect of the interphase differences and observed that in the homophase conditions, the signal was reported as coming from the center of the head, while in the antiphase conditions, the signal was reported as lateralized to the ears. Webster (1951) reported the same change in signal location in his experiment with masking thresholds. Hirsh further observed that the perceived change in signal location was a result of additional timing
and intensity cues that occur in the antiphase condition due to the out-of-phase signals. Signals that are in-phase, by contrast, are homogeneous in timing cues and, therefore, more difficult to hear, resulting in higher masked thresholds. This is consistent with the general agreement that the human sensory system identifies change more readily than static sensory stimulation. Thus, it is easier to hear the signal in $N_oS_π$ conditions than in either $N_oS_0$ or $N_πS_π$.

One of the main goals of APD testing has been to identify areas of the CNS that may be damaged or dysfunctional in the auditory processing pathways and surrounding structures. Binaural interaction tests primarily assess brainstem function (Johnson et al., 2007); however most behavioral binaural interaction tests are only sensitive to severe brainstem dysfunction and, as such, are not particularly clinically useful (Bellis, 2003, Johnson et al., 2007). As an exception to this, the MLD has been found to be sensitive to less extreme dysfunction (Noffsinger, Schaefer, & Martinez 1984), making the MLD one of the few binaural interaction tests that are appropriate for clinical APD assessment.

The use of MLDs on APD populations has been limited by the lack of a gold standard (ASHA, 2005; Bellis, 2003). However, the MLD test has been administered to various populations with cortical dysfunction, which has given a general indication of characteristic test performance for those with cortical disorders. In testing MLD, Olson and Noffsinger (1976) discovered that patients with multiple sclerosis and unilateral menier’s disease both exhibited smaller than normal MLDs. Patients with lower brainstem lesions, particularly in the pontomedullary junction, exhibit significantly smaller or non-existent MLDs (Lynn, et al., 1981). In addition, patients with cochlear pathology also exhibit significantly smaller MLDs (Quaranta & Cervellera, 1974). The indication from test findings shows consistently that cortical dysfunction leads to small or non-existent MLDs. Thus, it is logical to expect that abnormal
central processing function would also result in small or non-existent MLDs. This is consistent with the idea that individuals with APD have a more difficult time processing the spatial and temporal cues that should make it easier to identify an antiphasic signal. Thus, individuals with APD do not exhibit as great a difference between antiphasic and homophasic thresholds as normal individuals do.

The MLD is a test that provides information on binaural interaction regarding temporal and spatial processing ability and can indicate dysfunction below the cortex in the brainstem area of the CNS. MLDs are an important component of test batteries for APD. Specifically, tonal MLDs provide a useful tool in differentiating between normal and non-normal patients. Masking noise having a center frequency equal to the signal frequency is presented in both homophasic and antiphasic conditions. MLDs that are significantly lower than normal are an indication of auditory processing dysfunction and indicate further APD testing.

**Method**

**Present Study**

To fill the need for more effective diagnostic measures of APD, the Institute of Physiology and Pathology of Hearing (Warsaw, Poland) and the Brigham Young University Communication Disorders Department (Provo, Utah) have collaborated to produce a software program designed to measure several different central auditory processes using shortened and adapted psychometric test procedures in an effort to make the tests more clinically functional. The software is still under development but currently includes adapted versions of the following tests: difference limens for frequency, intensity and duration, gap detection test, temporal order threshold, interaural time difference, interaural intensity difference, masking level difference,
pattern recognition tests for frequency and duration, dichotic digits test, competing word sequence test, speech in noise, compressed word test, and filtered word test.

Collectively, these tests will provide measurable data for each of the central auditory processing skills identified by the ASHA (2005) technical report as important components of APDs. In a clinical setting, it would be rare for all of these tests to be administered to the same patient since APD is so variable and audiologists are encouraged to adapt each patient’s diagnostic exam according to their patient’s individual difficulties (ASHA, 2005; Bellis, 2003; Chermak & Musiek, 1997). However, having all these tests available within one easily administered test battery would have great clinical value.

The present study addressed the need for more effective testing materials for APD diagnosis by administering 12 of the available tests from the developed software. However, the scope of this paper is such that only the MLD portions of the test process will be discussed.

Participants

Participants in this study included 20 males and 20 females between the ages of 18 and 26 years. All participants demonstrated normal hearing bilaterally with pure-tone thresholds ≤15 dB HL for octave intervals between 250-8000 Hz. Word recognition scores were ≥ 98% bilaterally. All participants had normal type A tympanograms and otoscopic examinations revealed clear ear canals and intact eardrums, bilaterally. All participants completed an interview (Appendix A) reporting they were native English speakers with a history free from head trauma, drug abuse, and neuropsychiatric disorders including but not limited to ADHD, cognitive delays, and learning impairments. Participants were also screened for handedness since hand dominance is highly correlated with hemispheric dominance. Most individuals are right handed and left hemisphere dominant. Those with right hemisphere dominance have been noted
to process information differently. Currently there is no non-invasive way to determine hemisphere dominance, but studies show that up to 92.5% of right-handed individuals are left hemisphere dominant (Knecht et al, 2000). Only right handed individuals were used in this study. All participants signed an informed consent document approved by the Institutional Review Board at Brigham Young University (Appendix B) prior to testing.

**Apparatus**

All testing was performed in a single walled, sound-isolated booth that met American National Standards Institute (ANSI) maximum permissible levels for ambient noise with ears uncovered and all electronic equipment operating (ANSI, 1999). The stimulus was presented using TDH-50P Telephonics earphones and a Grason-Stradler model 1761 audiometer. The audiometer was calibrated before and after the experiment according to ANSI standards (ANSI, 2004). A Grason-Stradler model GSI-33 Impedance Bridge was used to perform tympanograms.

Tests were administered using a Dell computer running Microsoft Windows 2007. All testing was administered in the single walled sound-isolated booth, with the output of the computer directed through the 1761 audiometer, as specified above. Calibration was completed according to the software design specification, prior to the first test for each new participant throughout the testing period.

**Procedures**

The specialized software used two different approaches for determining MLDs. The first method was referred to as an adaptive MLD (MLDA) and the second as Bekesy MLD (MLDB). To control for test order effects, the test order was randomized for each participant. The MLDA consisted of a two subtest conditions (N,So and N,SP); the MLDB consisted of four subtest conditions (N,oSo, N,oSP, N,pSo, and N,pSP). The order of the subtest conditions was also
randomized to control for test order effects. The stimulus for each subtest was presented according to the presentation levels and specifications of the specialized software and as detailed under the specific test headings below.

The stimuli consisted of noise and a pure-tone signal presented binaurally. The same noise and signal was used for both MLDA and MLDB, though the method of presentation differed. The noise was created according to the specifications outlined by Harris et al. (1992); the noise was filtered-random with a 600 Hz bandwidth (3-dB down) and centered at 500 Hz. The slope of the upper skirt was 44 dB/octave and the slope of the lower skirt was 40.6 dB/octave. The signal was a 500 Hz sinusoid tone of 500 ms duration.

**Adaptive Masking Level Difference.** The MLDA test used an adapted procedure to determine thresholds for two conditions: $N_oS_o$ and $N_oS_n$. The threshold for each condition was determined separately in two subtests. The target stimulus for each condition was a 200 ms presentation of noise with the signal embedded in the middle. The non-target stimulus was the noise without the signal. The participants were asked to press a button only when they heard the target stimulus. Subject instructions were as follows:

You will hear some noise in your ear. Sometimes there will be a tone with the noise and sometimes there will not be a tone. You are to press the button when you think you hear the tone. At first it will be very obvious if there is a tone or not, but as the test continues, the tone will become fainter. You are to press the button even if the tone is very faint. If you are not sure you hear the tone, guess. Are there any questions? OK, we will start the test now.

The test began with a training session to familiarize the participant to the task. The target and non-target stimuli were presented to the participant. The participant was then asked if
they would like to hear either stimulus again. This was repeated until the participant indicated she or he was ready to begin testing.

The threshold was determined in a two-step procedure. In step one, only the target stimulus was presented. Initially, the stimulus was presented with the noise at 50 dB HL and the signal at 30 dB HL. The signal intensity decrease by half each time the participant correctly identified the stimuli (hit). This process continued until the participant could no longer identify the stimuli, which was indicated by the participant not pressing the button (miss). The signal intensity was then increased by half of the previous increment. The signal intensity then decreased by half until the next miss. At that point, step one ended and step two automatically began.

In step two, the participant was presented with both target and non-target stimuli in random order. The initial intensity was determined by the lowest hit value plus 2 dB. In addition to hits and misses, the software recorded when the non-target was correctly identified (correct rejection) and when the non-target was incorrectly identified (false alarm). After each hit, the signal intensity decreased by 2 dB. After each miss or false alarm, the signal intensity increased in 2 dB increments. A correct rejection resulted in no change to the signal intensity. Each change in direction of intensity movement (increasing or decreasing) signified a reversal. Low reversals were identified as the lowest intensity at which the participant heard the tone before generating a miss. The test continued until seven reversals were completed. The software then calculated the threshold for each condition by averaging the target values for the five best low reversals.

The software recorded the thresholds for \((N_oS_o)\) and \((N_oS_a)\) separately. Hits, misses, correct rejections and false alarms were plotted on a graph. Using signal detection theory, the
software also plotted the Receiver Operating Characteristic curve (ROC) and calculated the
discriminability index ($d'$).

**Bekesy Masking Level Difference.** The second approach used to determine MLD was a search for the listener’s threshold where the participant responded to a tone that varied in intensity using a Bekesy tracking method. This was referred to as MLDB. Masking levels were generated for four conditions: noise and signal in-phase ($N_oS_o$) noise in phase, signal 180° out-of-phase ($N_oS_{\pi}$), noise 180° out-of-phase, signal in-phase ($N_{\pi}S_o$), and both noise and signal out-of-phase ($N_{\pi}S_{\pi}$).

The stimulus presented was significantly different to that of the MLDA test. For MLDB, the stimulus consisted of a constant noise and a repeating 500 Hz signal. The 500 Hz sinusoid occurred at a rate of one interruption per second. The participants were required to press and continue to hold down a button for as long as they could hear the repeating tone, and then release the button when the tone was not heard. Participant instructions were as follows:

You will hear some noise in your ear. At the same time you will hear a tone presented with the noise. When you press the button and hold it down, the tone will become quieter and quieter. You are to press the button and hold it down until you are sure you cannot hear the tone any longer. When you release the button, the tone will start to get louder. As soon as you think you hear the tone again, press and hold down the button. Continue this process for as long as you think you hear the tone. If you are not sure if you hear the tone, guess. Are there any questions? OK, we will start the test now.

According to the software specifications, the stimulus was first presented with the noise at 40 dB HL and the signal at 30 dB HL. As the participant held down the button, the noise remained constant while the signal attenuated at a rate of 2.5 dB/s. When the participant released
the button, the signal increased in intensity at the same rate. The software graphed the responses for each stimuli and recorded thresholds and total test duration time. Each change between increasing and decreasing intensity was an excursion and was bound by peak and trough values. A threshold was determined for each excursion by summing the peak and trough values and dividing by 2. The test continued until 15 excursions were completed. The threshold was then determined by calculating the mean of the midpoints for the 10 best excursions. The software calculated three masking differences, as follows: NoSo - NoSπ; NoSo - NπSo; NoSo - NπSπ.

Results

Descriptive Data

Descriptive statistic measures were computed for both the MLDA and MLDB procedures. Threshold values and MLDs (NoSo - NπSπ) are shown separately for the two procedures in Table 1. It should be noted that under both procedures, the mean threshold for NoSo was higher than the mean threshold for NπSπ. Though variability in MLD between the two procedures is comparable, MLD is much lower in the Bekesy procedure (9.15 dB) than in the adaptive procedure (15.33 dB).

Inferential Data

Combined Procedures. A MANOVA of the combined results for both procedures showed a significant difference for MLD in both procedure and sex. The MDLA and MLDB procedure difference was significant; F(3, 74) = 32.701, p ≤ 0.001. However, the interaction between procedure and sex failed to show significance; F(3, 74) = 1.401, p ≤ 0.249. A between-subject analysis showed a significant difference between procedures for NoSo at F(1, 38) = 7.155, p ≤ 0.009, and for NπSπ at F(1, 38) = 45.612, p ≤ 0.001. The analysis also showed a significant difference between procedures for MLD at F(1, 38) = 89.550, p ≤ 0.001.
Table 1

*Average thresholds and MLDs for Adaptive (MLDA) and Bekesy (MLDB) procedures (female = 20, male = 20, combined = 40)*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sex</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLDA Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_0S_o$ (dB HL)</td>
<td>F</td>
<td>21.90</td>
<td>1.96</td>
<td>17.40</td>
<td>27.40</td>
<td>20.60</td>
<td>23.40</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>20.57</td>
<td>2.85</td>
<td>10.40</td>
<td>24.00</td>
<td>18.68</td>
<td>22.26</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>21.24</td>
<td>2.50</td>
<td>10.40</td>
<td>27.40</td>
<td>20.43</td>
<td>22.04</td>
</tr>
<tr>
<td>$N_0S_\pi$ (dB HL)</td>
<td>F</td>
<td>6.64</td>
<td>2.44</td>
<td>2.20</td>
<td>10.80</td>
<td>4.07</td>
<td>8.37</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>5.20</td>
<td>1.75</td>
<td>2.20</td>
<td>8.40</td>
<td>3.70</td>
<td>5.82</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>5.91</td>
<td>2.22</td>
<td>2.20</td>
<td>10.80</td>
<td>5.20</td>
<td>6.62</td>
</tr>
<tr>
<td>MLD (dB)</td>
<td>F</td>
<td>15.26</td>
<td>2.74</td>
<td>10.40</td>
<td>20.60</td>
<td>13.35</td>
<td>16.90</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>15.39</td>
<td>3.15</td>
<td>4.40</td>
<td>19.80</td>
<td>13.20</td>
<td>16.91</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>15.33</td>
<td>2.91</td>
<td>4.40</td>
<td>20.60</td>
<td>13.88</td>
<td>16.29</td>
</tr>
<tr>
<td>MLDB Procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_0S_o$ (dB HL)</td>
<td>F</td>
<td>20.53</td>
<td>4.01</td>
<td>10.50</td>
<td>25.40</td>
<td>18.65</td>
<td>22.41</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>18.16</td>
<td>3.47</td>
<td>10.40</td>
<td>22.70</td>
<td>16.53</td>
<td>19.78</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>19.34</td>
<td>3.89</td>
<td>10.40</td>
<td>25.40</td>
<td>18.10</td>
<td>20.59</td>
</tr>
<tr>
<td>$N_0S_\pi$ (dB HL)</td>
<td>F</td>
<td>11.82</td>
<td>3.36</td>
<td>1.80</td>
<td>15.60</td>
<td>10.25</td>
<td>13.39</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>8.55</td>
<td>3.43</td>
<td>1.70</td>
<td>14.00</td>
<td>6.94</td>
<td>10.15</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>10.18</td>
<td>3.74</td>
<td>1.70</td>
<td>15.60</td>
<td>8.99</td>
<td>11.38</td>
</tr>
<tr>
<td>MLD (dB)</td>
<td>F</td>
<td>8.72</td>
<td>3.20</td>
<td>-0.50</td>
<td>13.50</td>
<td>7.22</td>
<td>10.22</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>9.60</td>
<td>2.52</td>
<td>4.50</td>
<td>13.00</td>
<td>8.42</td>
<td>10.77</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>9.15</td>
<td>2.88</td>
<td>-0.50</td>
<td>13.50</td>
<td>8.24</td>
<td>10.08</td>
</tr>
</tbody>
</table>

*Note. CI = confidence interval. F = female, M = male, C = combined female and male.*
The MANOVA for female and male sex difference was significant at $F(3, 74) = 5.608$, $p \leq 0.002$. A between-subject analysis showed a significant difference between female and male for $N_oS_o$ at $F(1, 38) = 6.855$, $p \leq 0.011$, and for $N_oS_\pi$ at $F(1, 38) = 14.005$, $p \leq 0.001$. The MANOVA failed to show a significant difference between female and male for MLD; $F(3, 74) = 5.050$, $p \leq 0.443$. Likewise the MANOVA failed to show a significant interaction between procedure and sex for $N_oS_o$ at $F(1, 38) = 0.545$, $p \leq 0.462$, for $N_oS_\pi$ at $F(1, 38) = 2.058$, $p \leq 0.156$, and MLD at $F(3, 74) = 0.327$, $p \leq 0.569$. Further analysis was completed separately for both MLDA and MLDB to determine the source of the significant sex differences

**Adaptive Masking Level Difference.** A one-way ANOVA was completed for both the $N_oS_o$ and $N_oS_\pi$ conditions and the MLD value by sex. The result showed a significant difference between female and male results for the $N_oS_\pi$ condition at $F(1, 38) = 4.735$, $p \leq 0.036$. No significant difference was found between female and male for either the $N_oS_o$ condition at $F(1, 38) = 2.962$, $p \leq 0.093$, the $N_oS_\pi$ or for the MLD at $F(1, 38) = 0.19$, $p \leq 0.890$.

**Bekesy Masking Level Difference.** Analysis using a one-way ANOVA showed some significant differences between female and male for masked thresholds: The difference between female and male for $N_oS_\pi$ was significant at $F(1, 38) = 9.321$, $p \leq 0.004$ as well as between female and male for $N_oS_o$ at $F(1, 38) = 4.533$, $p \leq 0.040$. No significant difference was found between female and male thresholds for either $N_oS_o$ at $F(1, 38) = 4.014$, $p \leq 0.052$, $N_oS_\pi$ at $F(1, 38) = 3.957$, $p \leq 0.054$, or for MLD at $F(1, 38) = 0.924$, $p \leq 0.343$.

**Discriminability Index**

For MLDA, hits and misses were plotted and used to illustrate the participants’ ROC curve and determine the discriminability index as measured by $d'$. Descriptive statistics on average $d'$ values are shown in Table 2. The $d'$ was not calculated for 12 participants in the
condition or for 20 participants in the NoSπ condition. This is because d’ is calculated by plotting hits vs. false alarms. Thus, for participants who had 100% hits with zero false alarms (or zero hits with 100% false alarms), d’ could not be calculated.

Means for each condition (NoSo and NoSπ) were similar. Variability, as measured by the standard deviation, was also similar for each condition, with the lowest standard deviation for females in the NoSπ condition and the highest standard deviation for males in the NoSo condition. It is of note that d’ was computed more frequently in the NoSo condition (N = 28) than in the NoSπ condition (N = 20), and more often for males (N = 26) than for females (N = 22).

Table 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Sex</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>95% CI Low</th>
<th>95% CI High</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoSo</td>
<td>F</td>
<td>13</td>
<td>1.35</td>
<td>0.35</td>
<td>0.73</td>
<td>1.82</td>
<td>1.14</td>
<td>1.55</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>15</td>
<td>1.31</td>
<td>0.40</td>
<td>0.44</td>
<td>1.87</td>
<td>1.10</td>
<td>1.53</td>
</tr>
<tr>
<td>NoSπ</td>
<td>F</td>
<td>9</td>
<td>1.47</td>
<td>0.25</td>
<td>0.97</td>
<td>1.90</td>
<td>1.27</td>
<td>1.66</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>11</td>
<td>1.53</td>
<td>0.30</td>
<td>1.00</td>
<td>1.90</td>
<td>1.33</td>
<td>1.73</td>
</tr>
</tbody>
</table>

Note. CI = confidence interval. M = male, F = female.

A one-way ANOVA failed to show a significant difference in d’ for female and male participants in the MLDA procedure at $F(1, 46) = 0.010, p \leq 0.922$. The combined female and male average for d’ was 1.4 ($SD = 0.34$). MLDA test sensitivity was found to be 96.4% with specificity at 60.3%. This is shown in Figure 1 where ROC curves are depicted with an area of 0.87 under the ROC curve.
Figure 1. Receiver Operating Characteristic (ROC) curve for the combined female and male participants for the MLDA Test. The grey lines indicate $d'$ parameters for $d'$ values of 0, 1, 2, 3, and 4. The black line indicates the average $d'$ for the MLDA and the dotted lines indicate $+1$ SD for the MLDA. The computed $d'$ for the MLDA procedure was 1.4.

Discussion

In the present study, the data showed consistently higher thresholds for $N_0S_0$ than for $N_0S_π$. This observation is consistent with data obtained from other studies (Hirsh, 1948; Licklider, 1948; Olson, 1976; Wilson, 1984). Hirsh (1948) and Yost (2007) stated that the difference in threshold between $N_0S_0$ and $N_0S_π$ is due to the fact that antiphasic conditions provide phase cues from differences in localization of in-phase and out-of-phase stimuli which homophasic conditions lack. The additional phase cues of the $N_0S_π$ antiphasic condition results in a lower threshold, most likely due to the phase reversal triggering an alerting response in the sensory system.
Although females in this study had higher mean thresholds than males for all conditions, statistically significant differences were only found for the $N_0S_{\pi}$ condition. This observation was true for both the MLDA and MLDB procedures. However, it is of note that despite the significant difference in thresholds, the results failed to show a significant difference between female and male for the MLD. This indicates that although individual threshold values may vary between sexes, the actual MLD value is similar and, therefore, normative MLD data does not need to be reported separately for the two sexes.

Statistically significant differences in this study were also found between procedures, showing higher threshold values for $N_0S_{\pi}$ using the MLDB procedure than for the MLDA. Because of the higher $N_0S_{\pi}$ thresholds in MLDB, MLD values were lower in the MLDB procedure ($M = 9.15$ dB) than for the MLDA procedure ($M = 15.33$ dB). The greater difference between $N_0S_0$ and $N_0S_{\pi}$ in MLDA shows an increased differentiation between conditions in the MLDA procedure and, thus, suggests that MLDA may be the more useful procedure.

A possible reason for the difference between MLDA and MLDB is the difference in response time. In a Bekesy procedure, the stimulus is constantly changing at a set rate of 2.5 dB/sec. In the time it takes the participant to make a judgment on the presence or absence of stimuli and then produce the requested response (either push or release a button), the stimulus has continued to either increase or decrease. This results in an overshooting of the actual threshold. The Bekesy procedure attempts to compensate for this by using the midpoints rather than peaks or troughs to calculate the MLD, however the process is not, by its very nature, very precise.

In contrast, the MLDA procedure presents single discreet trials of the stimulus. The stimulus intensity changes in 2 dB increments but there is no interstimulus interval. Thus, the
patient has a longer time to determine if the stimulus was present or absent. This method also
has the added advantage of recording not just threshold levels but hits, misses, false alarms, and
correct rejections. Using this information, it is possible to calculate $d'$ and the ROC curve, which
gives information on test sensitivity and response bias. No similar value can be computed for the
Bekesy procedure since it does not calculate hits and misses. Thus, we are unable to report on
the test sensitivity of MLDB. It seems probable that MLDA is a more precise and valid
procedure because, unlike the MLDB procedure, the MLDA provides sensitivity and specificity
information and also allows the participant time to process and respond to discrete trials at
specific intensity levels.

Traditionally, MLD is most often calculated using a Bekesy method. This is likely due to
the quickness with which thresholds are obtained using the Bekesy method. In our study, it took
an average of 1.25 minutes per condition to obtain thresholds using MLDA. By contrast, it took
an average of only 0.95 minutes per condition to obtain thresholds using MLDB. However, the
time advantage of the MLDB procedure is less important to the overall test structure than the
sensitivity and specificity index and the additional processing time allowed in the MLDA
procedure, indicating that although MLDA takes longer to administer, it may be the better
clinical testing procedure.

In order to better compare MLD values between MLDA, MLDB, and the values obtained
by other studies, a meta analysis (Borenstein, Hedges, Higgins, & Rothstein, 2009) was
completed using the present study, Harris et al. (1992), Jerger et al. (1984), Wilson, et al. (1984),
and Olson et al. (1976). These different experiments varied slightly, but all had the same
attenuation rate of 2.5 dB/sec and used the same type of frequency modulated noise. Thus, it
follows that the results should be somewhat comparable. The meta-analysis used a fixed-effect
model to evaluate across groups. The standard error of the mean for each study was computed by dividing the sample standard deviation by the square root of the sample size (Schiavetti & Metz, 2006).

The results of the meta-analysis showed a Q-value of 105.27 that was significant at \( p \leq 0.01 \). This showed that there was greater variance across studies than would have been expected from a within-study error (Bornstein et al., 2009; p. 191). This observation remained true even when one or both of the MLD procedures used in the present study were removed from the analysis, although in this condition the Q-value decreased to 12.21 with a significance of \( p \leq 0.002 \). These results suggest that variability across studies is an ongoing difficulty in MLD testing. This is corroborated by the Z-value (135.28) for the fixed-effect models, which was significant for \( p \leq 0.01 \), again showing a greater variability than would be expected from the within-study error. These results clearly indicate that, due to the variability within tests of MLD, it is necessary to individually establish normative data for each clinical test protocol. Therefore, a standardized computer controlled procedure for obtaining MLD has particular significance since it would allow those using the same procedure to have access to previously established normative data and compare across facilities.

The analysis of d’ indicated that the MLDA had good test sensitivity (96.4%) but poorer specificity (60.3%). This suggests that the MLDA procedure was better at identifying those with APD than at identifying those who do not have APD. It is interesting to note that the analysis failed to identify differences in d’ between males and females. This suggests that the test itself does not exhibit sex-bias and, therefore, significant sex-differences between thresholds in the various conditions could be due to actual differences in female and male performance.
The reduced number of d’ subjects brings up another point of interest. For a participant to complete the test with either 100% hits or 100% false alarms indicates that the participant is being either too cautious, or does not fully understand the task. The value $d' = 1.4$ was calculated using data from both MLDA conditions ($N_oS_o$ and $N_oS_n$) and indicates typical response bias for MLDA and is an indication of good test construct. However, this d’ value only reflects 60% of the test since d’ was incalculable for the remaining 40%. For the participants for whom $d'$ was unable to be calculated it is arguable that no true threshold values were obtained. 

Audiological thresholds are determined by finding the point where the participants can hear a signal 50% of the time. Thus, participants with hit rates of 100% or zero have not presented a true threshold. In a clinical setting when d’ is not calculated, it would be advisable to reinstruct the participant, perhaps giving more encouragement to guess when they are not sure, and then re-administer the test.

Of the two tests analyzed, it appears that the MLDA is a more valid measure of MLD. Not only are the thresholds more comparable with those of previous literature, but the MLDA test also provides important test sensitivity and specificity information, using d’ and the ROC curve. It is encouraging to see that the data failed to show significant sex-difference in the MLD calculations, though it may be interesting for future studies to investigate the apparent sex-differences found in masked thresholds. The participants for whom no d’ was calculated raise important questions about overall test structure and also warrants further investigation. Future studies of the software developed by the Institute of Physiology and Pathology of Hearing and the Brigham Young University Communication Disorders Department should also focus on collecting normative samples from children and, if possible, from populations with APD. The results in this study indicate that a clinical software program with clear normative data that can
be compared across facilities would greatly improve the validity and ease of administration for current APD test batteries.


doi:10.1126/science.161.3837.184


Appendix A

Interview Questions

Pre-testing Interview Questions to Determine Participant’s Eligibility

David L. McPherson, Ph.D.
Department of Communication Disorders
Brigham Young University
(801) 422-6458

Name of Participant:__________________________ Date of Birth:______________

Purpose of the Interview
The following will be read to each prospective participant prior to the hearing screening or tests of Auditory Processing Disorder. These questions provide important information used to screen for ineligible participants. The participant’s answers must match the bolded response option for each of the six questions for the participant to be considered eligible to continue in the study.

Instructions to the Participant
I will read a series of interview questions. Please answer each question to the best of your knowledge. You may request clarification at any time during the interview process. Understand that you participation is completely voluntary and you are not obligated to answer the questions; at any point of the interview process, you may choose not to reply and to withdraw from the study.

Interview Questions
1. Are you right handed? (YES / No)
2. Are you a native English speaker? (YES / No)
3. Have you ever been diagnosed with a hearing loss of any type? (Yes / NO)
4. To the best of your knowledge, are you currently experiencing any ear infections, broken ear drums or any other hearing abnormalities? (Yes / NO)
5. Do you have a history of head trauma or drug abuse? (Yes / NO)
6. Do you have a history of neuropsychiatric disorders which include, but are not limited to, Attention Deficit and Hyperactive Disorder (ADHD), cognitive delays or learning impairments? (Yes / NO)
Appendix B

Informed Consent

Informed Consent to Act as a Human Research Subject

David L. McPherson, Ph.D.
Department of Communication Disorders
Brigham Young University
(801) 422-6458

Name of Participant: _________________________  Date of Birth: ____________

Introduction

This research study is being conducted at Brigham Young University by Professor McPherson, PhD, and graduate student Maria Burnham. The study examines individual responses to sound stimuli (such as clicking a button in response to a sound) in normal hearing individuals and will provide important data for future assessment of individuals with Auditory Processing Disorder (APD).

Procedures

This study will be conducted in room 111 of the John Taylor Building on the Brigham Young University campus. You will be asked to complete a preliminary interview, hearing screening and eleven APD tests. The interview includes six short yes/no questions regarding your English proficiency and hearing ability as well as any history of head injury, drug use, learning impairments or neuropsychological difficulties. The hearing screening includes the following hearing tests: Pure-tone Audiometry, Speech Recognition Thresholds, Tympanometry and Otoscopy. The eleven APD tests will require you to listen to several different kinds of sound stimuli and respond as instructed by the investigator. The interview, hearing screening and APD testing will all be conducted in one session which will take approximately 90 minutes.

Risks/Discomforts

Participation in this study is free of charge and does not include any treatment or intervention procedures; there are no known risks associated with the testing procedures. However, it is possible that the earphones used during the hearing screening and testing process may cause you to feel some discomfort. If this should occur, please inform the Investigator immediately; you will be given a short break and the earphones will be adjusted.

Benefits

There are no direct benefits to the subjects. However, it is hoped that your participation will provide important data needed to standardized and improve the efficiency of current APD tests.

Confidentiality

All information obtained from before and during testing procedures is strictly confidential and protected by governing privacy laws. Information specifically pertaining to you will not be released without your signature. All collected test results will be reported without
identifying information. All identifying references will be removed and replaced with control numbers. All data, including interview questions and screening/test results will be kept in a secure area that is only accessibly to those directly associated with the study. At the end of the study, all identifying materials will be destroyed. In addition, if at any time during the study you become ineligible or choose to discontinue, all information and data gathered from you will be shredded.

**Compensation**
All participants will receive a $5 gift certificate to Little Caesar’s. You will receive this compensation regardless of test performance or screening results. Your gift certificate will be given to you at the conclusion of the session.

**Participation**
Participation in this research study is voluntary. You have the right to withdraw at anytime or refuse to participate entirely without jeopardy to your class status or university standing.

**Questions about the Research**
If you have questions or concerns regarding this study, you may ask the investigators at any time during the session or contact Maria Burnham, Graduate Student, Department of Communication Disorders, 111 Taylor Building, Provo, Utah 84604; phone (360) 920-0547; email marianoelle.b@gmail.com; or you may contact David McPherson, Ph.D., Department of Communication Disorders, 129 Taylor Building, Provo, Utah 84602; phone (801) 422-6458; email: david_mcperson@byu.edu.

**Questions about your Rights as Research Participants**
If you have questions regarding your rights as a participant in this research project, you may contact Christopher Dromey, PhD, IRB Chair, 133 TLRB, Brigham Young University, Provo, Utah 84602; phone (801) 422-6461; email: Christopher_Dromey@byu.edu.

I have read, understood and received a copy of the above consent form. The procedures have been explained to my satisfaction and questions relating to procedural risks have been answered. I choose of my own free will to participate in the study explained above.

________________________  ____________________
Signature of Participant     Date