# VISUAL REINFORCEMENT INFANT SPEECH DISCRIMINATION: DEVELOPING A METHOD OF PERFORMANCE ANALYSIS

by

Tammy Fredrickson

B.S., University of Wisconsin, 1996

M.S., University of Wisconsin, 1998

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Speech, Language, and Hearing Sciences

2010





ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis entitled:

Visual Reinforcement Infant Speech Discrimination: Developing a Method of Performance Analysis

Written by Tammy L. Fredrickson

has been approved for the Department of Speech, Language and Hearing Sciences

_____
Christie Yoshinaga-Itano, Ph.D.

_____
Phillip M. Gilley, Ph.D.

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

HRC protocol #0906.8
COMIRB #04-0080

Fredrickson, Tammy L. (Ph.D., Speech, Language and Hearing Sciences)
Visual Reinforcement Infant Speech Discrimination: Developing a Method of Performance Analysis

Thesis directed by Christine Yoshinaga-Itano and Phillip M. Gilley

While measures of speech perception are an important aspect of audiological assessment and validation of amplification fitting in older children and adults, no clinical method of speech perception assessment exists for infants and toddlers. Visual Reinforcement Infant Speech Discrimination (VRISD) is a tool that has been used to assess infant speech perception in studies for over 30 years and has been deemed to have potential for clinical use. Unfortunately, the foundational work to provide an appropriate protocol for VRISD's clinical use does not exist; the reliability and validity of VRISD have not been studied. In its current research form, VRISD consists of 30 trials and is designed not as an assessment of an individual child's ability to discriminate specific phoneme or consonant vowel comparisons, but as a technique to compare the mean performance of a group of children with another – by age, by native language, or to determine whether one comparison is more difficult than another. This dissertation aimed to begin investigation into the foundational work necessary to help VRISD become a clinical tool. VRISD was used to assess 15 normal-hearing infants' abilities to discriminate three speech sound contrasts (a/i, ba/pa, and da/ga). Results of

infants' performance on these three contrasts indicated a hierarchy of difficulty, with the a/i contrast being the easiest and the da/ga contrast being the most difficult, as hypothesized. Results from each of the contrasts tested were analyzed in six different methods – five of which have been used in previously published VRISD studies – so that conclusions made from the different methods could be compared. The different methods of analysis sometimes led to differing conclusions as to the ability of an infant to discriminate a contrast. It was determined that the use of a criterion based on binomial probability was the best way to analyze performance in a manner that would yield reliable results as well as provide construct validity.

The current study revealed that individual children have performance profiles that indicate that they do not perform consistently across 30 trials, particularly when they demonstrate mastery or correct performance consistently over the first 10-12 trials. This inconsistent performance has previously resulted in a significant number of children who have been unable to complete the task or whose performance is correct in the beginning of testing and then becomes incorrect, possibly because of habituation or boredom. There has previously been no universally accepted and well-defined criterion for establishing whether an individual child has mastered a specific discrimination task. This dissertation compares six different methods of determining VRISD performance. A criterion-based performance measure using binomial probability provides the best technique for construct validity.

Acknowledgements

While progress with VRISD research has sometimes been more difficult that I'd anticipated, it's really exciting to think about its potential implications.

Many thanks to my committee members: Lew Harvey, Don Finan, Rafael Delgado, Anu Sharma, and especially my co-advisors Christie Yoshinaga-Itano and Phillip Gilley. Thank you to David Henry, Roanne Karzon, and Gerald Popelka, all of whom served as great mentors along the way and encouraged me to continue my studies.

Special thanks to Rafael Delgado and Intelligent Hearing Systems for creating our VRISD software and for being open to feedback and suggestions for future versions of VRISD. I look forward to continuing the development and utilization of VRISD!

Thank you to the folks who helped me by assisting with testing: Molly Dalpes, Lauren Johnson, Kristin Sommerfeldt, Megan Burgess, Melinda Anderson, Kristin Uhler, Stacy Claycomb, and Darcy Strong. The job of the assistant is the most important one – so thanks!

Thanks to my "VRISD-partner-in-crime" Kristin Uhler - I'm so thankful to have had you with me for this venture. Your ability to listen, brainstorm, and encourage me is much appreciated!

And thank you for support and encouragement from Holly Bishop, Darcy Strong, Jen Barnes, my mom Kathy, and numerous others who frequently checked in to see how research was progressing and were excited

about the possibilities of VRISD. Sometimes it was their excitement that kept me going!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1: INTRODUCTION

Universal newborn hearing screening has become commonplace in the United States and other areas of the world over the last 10 years. According to the National Center for Hearing Assessment and Management, 42 states (as well as the District of Columbia and Puerto Rico) currently have laws or statutes that make newborn hearing screening mandatory ("EHDI Legislation," 2009). The Centers for Disease Control and Prevention reports that 94% of newborns in the United States and its territories received hearing screenings in 2007 ("Early Hearing Detection and Intervention (EHDI) Program," 2010). Based on data reported, approximately 6% of those infants screened in 2007 were diagnosed with permanent hearing loss (1.2 per 1000 babies screened).

Those infants identified with permanent hearing loss are fit with amplification earlier than was possible prior to newborn hearing screening, with an average age of identification of 3 months or lower in many states ("Summary of Infants Diagnosed Before 3 Months of Age ( Year 2007)," 2009). While validation of the fitting of amplification in older children and adults can be made using both Real Ear measurements and speech perception measures, infants and young children are fit with amplification through prescriptive formulas and validation of fitting is evaluated through Real Ear measurements (typically Real Ear to Coupler (RECD) measurements) and observations of their behavior. There is no standard method of speech perception assessment for this population that has been

incorporated into the clinical battery of audiometric testing. Typically, infants' and young children's amplification benefit is assessed via methods such as clinical observation, behavioral observation audiometry, visual reinforcement audiometry and/or subjective parent questionnaires such as the Infant-Toddler Meaningful Auditory Integration Scale (IT-MAIS) (Zimmerman-Phillips, McConkey Robbins, & Observer, 2001; Zimmerman-Phillips, Osberger, & Robbins, 1997) and the Parents' Evaluation of Aural/Oral Performance of Children (PEACH) (Ching & Hill, 2007). These questionnaires are completed by the infant's parents and those who work closely with the infant. Unfortunately, these methods do not yield objective information regarding the speech perception abilities of these infants and none of these measures have standardized or validated results. These questionnaires cannot provide specific information regarding the infant's ability to discriminate one phoneme from another. Because these questionnaires address an infant's perception in his/her everyday environment, there is an inability to control auditory parameters of the stimuli, such as intensity, duration, speech versus non-speech, or to allow discrete comparisons of one auditory stimulus as compared to another.

Therefore, there is a critical need for the development of a standard measure for speech perception in infants. Currently, standardized speech perception tasks involve real words and therefore, cannot be used until the child has acquired language at levels appropriate for a typically developing two to three year old child. Objective assessment of an infant's ability to

discriminate speech when he/she has a significant hearing loss could yield additional information about the infant's overall perceptual development as well as the performance of his/her amplification. An objective measure of infant speech perception abilities could help to verify that an infant's amplification provides the sound necessary for the development of spoken language. Results obtained using infant speech perception measures could also help to direct habilitation efforts as well as monitor those efforts.

# CHAPTER 2: REVIEW OF THE LITERATURE

Speech perception skills in infants who have normal hearing and are typically developing have been studied in depth for decades. Most of this research has been done in the field of linguistics to assess the development of speech perception and infants' ability to detect particular acoustic cues, comparing these abilities with those of adults (e.g., Aslin, 1981; Bull, Eilers, & Oller, 1984; Eilers, Bull, Oller, & Lewis, 1984; Eilers, Morse, Gavin, & Oller, 1981).

Various methods have been used to assess speech perception abilities in normal hearing infants. Researchers have been interested in learning about the speech perception abilities of prelinguistic children as a way to learn about early language development and have compared the abilities of children and adults to determine when particular abilities develop. Methods used to assess speech perception in prelinguistic children include high amplitude sucking (HAS) (e.g.,Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Karzon, 1982), visual habituation (VH) (e.g., Bornstein & Benasich, 1986; Houston, Pisoni, Iler Kirk, Ying, & Miyamoto, 2003; Kaplan & Werner, 1986), Observer-Based Psychoacoustic Procedure (OPP) (e.g., Olsho, 1987), and Conditioned Head Turn (CHT), also referred to as Visual Reinforcement Infant Speech Discrimination (VRISD) (e.g., Eilers, Wilson, & Moore, 1977; Werker, Polka, & Pegg, 1997). These methods were primarily developed to allow investigators to examine early linguistic development, but they have also been used to measure thresholds. The methods were constructed to

permit investigation into the perceptual abilities of infants at different ages and stages of development. The following sections examine each of these techniques and provide examples of how they have been used in research (also see Jusczyk & Luce, 2002, for a review).

**High Amplitude Sucking –** The high amplitude sucking (HAS) habituation procedure (e.g., Eimas et al. 1971) is a method that can be used with very young infants, even newborns, to investigate babies' responses to changes in sound. With this method, babies hear a repeating speech sound while sucking on a pacifier that is connected electronically to a computer so that the rate of sucking behavior can be measured. Babies tend to start sucking at a high rate while hearing the repeating sound. As their rate of sucking decreases (i.e., indicating that the baby has habituated to the sound), infants in the experimental group will hear a new speech sound. The infants in the control group continue to hear the initial sound. If an increase in sucking rate is seen in the experimental group relative to the control group, it is said that discrimination has occurred. This is a demonstration of a reflexive response; no conditioning to the task is needed.

HAS has been used fairly successfully to provide valuable information about speech perception in very young infants. Data obtained via HAS is group data, therefore information about the abilities of individual infants cannot be obtained. HAS cannot address questions of audibility nor can it provide information as to how difficult a particular contrast may be for a particular infant. In addition, the attrition rate for the HAS procedure is usually

quite high, sometimes as high as 75% (e.g., Karzon, 1982). Also, typically no more than one speech sound contrast can be assessed per measurement period. Given these drawbacks, it is unlikely that HAS will become a part of a clinical battery to assess discrimination of speech in infants.

**Visual Habituation –** Visual habituation (VH) is a cognitive task that measures an infants' interest in sounds. This task has been used with infants ages 3 to 29 months. In a VH task, visual stimuli are presented via television screens or computer monitors. The monitors are placed on either side of the infant. Images such as checkerboards are displayed and are made to change in luminescence during a trial. This purpose of this change is to keep the infants' attention to the screens longer. Auditory stimuli are presented via a transducer such as a loudspeaker once the child begins fixating on a monitor. As long as the infant continues to look toward a monitor, the stimuli continue to play. The stimuli stop playing when the infant habituates to the sound and looks away from the monitor. The total amount of time the infant looks toward the monitor is recorded and is measured as an index of interest in the auditory stimuli. Familiarization trials occur first, during which time the infants' interest in the auditory stimuli decreases; the infant becomes habituated to the sound. Presentation with a novel stimulus, theoretically, results in prolongation of infants' visual attention toward the monitor. It is assumed that an infant will habituate to a frequent sound more quickly than he/she will habituate to a novel sound. Visual habituation assesses an infants' interest in novel stimuli by measuring the difference in looking time between the more

common stimulus and the novel one. Performance of an experimental group (babies who heard the novel stimuli) is compared with the performance of a control group (babies who hear only one stimulus), therefore, information regarding an individual infant's performance cannot be obtained. VH has been used to assess infants' linguistic development and ability to differentiate suprasegmental information (e.g., Houston et al., 2003; Spence & Moore, 2003).

This technique is a cognitive skill that measures an infant's interest in particular stimuli, not whether or not it can be heard and/or distinguished. VH is not necessarily assessing discrimination of speech sounds. It is important to note the attrition rate for VH tends to be variable and high, ranging from 17 to 62% (e.g., Miller & Eimas, 1996; Werker & McLeod, 1989). Methodological differences among the studies probably account for much of this variability in attrition rate. If VH is to be used in the clinical setting to assess prelinguistic cognitive perception, the attrition rate needs to reliably be less than 30% (as suggested for Visual Reinforcement Audiometry; Gravel, 1989). Also, in order to measure the duration of looking time, test sessions typically are video recorded and analysis of looking time and preference occurs after the test session is over. Without instant feedback regarding the infant's performance, changes in procedure that may have an immediate impact on performance cannot be implemented (i.e., re-training, adjusting stimulus intensity levels), which may result in a greater attrition rate. Clinical use of a tool that requires

analysis after the test session and has a high attrition rate is not practical in today's fast-paced clinical pediatric setting.

**Observer-based Psychoacoustic Procedure --** The Observer-based Psychoacoustic Procedure (OPP) (Olsho, Koch, Halpin, & Carter, 1987) was developed for use with infants 2 to 12 months of age. OPP combines features of behavioral observation audiometry (BOA) and visual reinforcement audiometry (VRA). OPP brings in the element of conditioning a response from the infant and, thus, turns this into a test of prelinguistic sensory perception. Typically, a conditioned response (i.e., a head-turn) cannot be elicited reliably until approximately 6 months of age. OPP permits investigations of auditory perception in younger infants that has otherwise not been possible. In this procedure, the infant sits on a parent's lap while sounds are presented through earphones to the infant. An assistant, who serves as a distracter by engaging the infant with toys, is seated on one side of the parent and infant. On the opposite side of the parent and infant is a mechanical toy visual reinforcer (the same as those used in visual reinforcement audiometry). An observer is seated in a control room and observes the infant in order to make judgments about whether or not a target sound was presented based on the infant's behaviors and responses.

When the infant is engaged with a toy, the observer initiates a trial. Two types of trials may occur: a signal trial or a no-signal trial. Signal trials are characterized by the presence of target stimuli. Target stimuli may include the onset of auditory stimuli or a change of auditory stimuli, depending on the

auditory skill being examined (i.e., detection or discrimination). No-signal trials are characterized by the absence of target stimuli (e.g. absence of sound or no change in stimuli). The observer, who is blinded to trial type by use of headphone sound attenuators or masking music, must decide if they believe the trial that has just occurred is indeed a signal or a no-signal trial based on the infant's behavior. If the observer correctly identifies that a signal trial has occurred, the toy reinforcer is activated. In contrast, if the observer incorrectly judges that a signal has occurred, the toy reinforcer is not activated. Results of each trial are displayed by the computer software, providing the observer immediate feedback.

For this procedure to be successful observers must undergo training, which was designed by the developers of this procedure to take approximately one month (Olsho et al., 1987). During training the observer becomes familiar with what types of responses constitute infant reactions to sounds and sound changes. These reactions may range from a head turn toward the stimulus to an eye blink or change in facial expression. Thus, OPP is similar to BOA in that it does not require that the child make a specific type of response (i.e., head turn), but allows for a variety of responses. However, OPP differs from BOA in that the observer is blinded to trial type and a toy reinforcer is used.

One limitation of OPP is a high false alarm rate. Behaviors such as an eye blink may be difficult to "read" as a response to a sound. Olsho et al. (1987) used a computer program that provided the observer feedback should

his/her false alarm rate rise above 25%. The observer then either stops the session entirely or begins retraining. After retraining, the observer's false alarm rate must not rise above 20% or testing is terminated. Olsho et al. (1987) reported 16% of sessions must be excluded due to a high false alarm rate on the part of the observer.

The ability to assess younger infants and thereby develop a better understanding of infant speech perception is a strong advantage of OPP. One benefit of OPP over VH for use in this young population is that OPP is a task of prelinguistic auditory perception rather than a prelinguistic cognitive task. OPP allows the examination of both auditory detection and discrimination.

Concerns raised when discussing the clinical utilization of OPP include observer training and its attrition rate. The length of time it takes to train observers (one month as suggested by Olsho et al., 1987) is typically not clinically feasible. Another disadvantage of OPP is its high attrition rate. A study conducted by Marean et al. (1992) investigated 2- and 3-month old infants' categorization of vowel sounds using the OPP procedure. Attrition rates for this study ranged from 29% (for 3 month old infants) to 53% (for 2 month old infants). This high attrition rate is a potential problem for the clinical usefulness of OPP, as it appears that "meaningful" results will be obtained on fewer than half of those infants who are tested. The attrition rates reported by Marean et al. were encountered in studies using normal hearing infants – one can anticipate an even higher attrition rate for children

with hearing loss. In a fast-paced clinical setting, this will result in a lot of time spent on a test that yields few results.

**Visual Reinforcement Infant Speech Discrimination –** Visual reinforcement infant speech discrimination (VRISD) is a variation of the commonly used VRA technique for assessing hearing sensitivity in infants and toddlers. Also known as the "conditioned head turn procedure" (CHT), VRISD was first developed by Eilers, Wilson, and Moore (1977) and modified in 1990 by Nozza. It has been proposed as being potentially useful in a clinical setting (Nozza, Miller, Rossman, & Bond, 1991). It has been used with normal hearing infants from 6 to 30 months of age with success, even in a clinical setting (Gravel, 1989). VRISD is a test of prelinguistic auditory perception, as is OPP, allowing for investigators to determine not only if a sound was heard, but also if it was distinguished from another sound. VRISD results have traditionally been analyzed as group data, analyzing differences in mean group performance on a particular task or contrast.

VRISD testing is based on an oddball paradigm where one speech sound serves as the background repeating stimuli, while another is presented periodically with the expectation that the child will detect this difference in sounds. One sound from each pair serves as the repeating background stimuli, while the other sound serves as the change or target stimuli. The child sits on a parent's lap or in a high chair in the center of the test room. An assistant, who serves as a distracter, sits directly in front of the child. Stimuli are delivered in the soundfield or through earphones. Trials are initiated by

the evaluator, who typically is in an adjacent sound booth and are initiated once the child's attention is directed toward the center. Trials are either control trials or target trials. During target trials, the novel stimulus is presented followed by a return to the background sound. In contrast, the background sound does not change during control trials. The evaluator, who is blinded as to whether the trial is a control or a target trial judges whether or not the child makes a head turn during the trial. Head turns are recorded on a computer and the computer determines if the child's response is a correct response or a false positive. The child is given visual reinforcement (i.e., the illumination of an animated toy or video) for correct responses.

Because VRISD requires a head turn response, it cannot be used with infants who do not have adequate head and neck control (normal developing infants tend to gain head and neck control around six months of age). VRISD, when successful, results in conditioned, reliable responses that can lend insight into an infants' ability to detect and discriminate speech sounds and its components (e.g., Eilers et al., 1984; Eilers et al., 1981; Nozza, 1987; Trainor & DesJardins, 2002). It is speculated that infants who can perform VRA testing can usually complete VRISD testing.

Werker, Polka, and Pegg (1997) summarized the history of VRISD as well as its strengths and weaknesses. VRISD has been used to assess infants' perception of speech, music, voices, and frequencies. Werker et al. reviewed both positives and negatives of VRISD testing, and concluded that VRISD is an "essential tool for the perceptual researcher and clinical

audiologist" (p.177, 1997). One limitation of the VRISD procedure is something that is seen again and again in infant speech perception test methods: a variable and fairly high attrition rate (rates as high as 76% (Bohn & Polka, 2001)).

There are several strengths of the procedure also explained by Werker et al. (1997). First, VRISD is a behavioral auditory test and, thus, is a test of functional hearing abilities. Second, the stimulus and reinforcer are independent of one another in this procedure, allowing for assessment of the infant's behavior/response to the stimuli separate from his/her response to the reinforcer. Third, multiple test trials can be presented, which permit the evaluator to determine if the infant can reliably discriminate specific stimuli. Finally, the procedure is flexible, permitting the use of various research designs.

**Summary –** Research performed using the various methods mentioned above have, with the exception of research performed by Houston using VH (Houston et al., 2003), been performed with normal hearing infants. Infants who are hard-of-hearing or deaf present a new challenge. Whereas with normal hearing infants, audibility of the signal is assumed, this assumption cannot be made with deaf or hard-of-hearing infants. Audibility can be assessed via aided auditory brainstem response testing or functional gain testing in the sound booth. However, audibility of the sound is only part of the process of learning speech and spoken language. Ability to discriminate sounds is necessary as well. Much research into infant auditory

discrimination has been attempted using auditory evoked potentials. This method holds much promise, but only tells us something about how the signal is perceived by the auditory pathway, not how the baby responds or hears behaviorally.

Most investigations using infant speech perception tests have not been attempted in the deaf and hard-of-hearing population. This is, in part, because the population of those infants identified with hearing loss early has historically been very small. In order to conduct investigations such as those discussed above, a method of assessment that is fairly easy and that most of this population of infants can indeed perform is required. Because infants are identified with hearing loss and intervention is begun very early in life, the need to assess perceptual abilities is great. A better understanding of the perceptual abilities of infants with hearing loss could lead to improvements in amplification and early intervention for these children.

VRISD appears to hold great potential for assessing hard-of-hearing infants' perceptual abilities. It is a test of functional hearing ability, which is what audiologists are interested in when fitting amplification for this population. Audiologists want to know how the infant hears with amplification in a practical sense – how the infant can hear and discern speech is vitally important. Since multiple test trials are possible even in only one sitting, the infant's ability to discriminate particular stimuli can be investigated. As noted above, its potential clinical use has already been acknowledged. The technique is easy to integrate into the clinic setting because it uses equipment

that is commonly used in infant hearing assessment: reinforcers (either mechanical toys or videos), loudspeakers, and a computer.

All infant speech perception test techniques have weaknesses in common: a fairly high attrition rate, stimuli must be fairly short in duration (therefore limiting the stimuli that can be used), babies do not reach 100% correct, and interpreting results (especially “negative” results) can be difficult. High attrition rates can make a tool such as VRISD difficult to incorporate into clinical practice. An experienced pediatric audiologist and an appropriate set-up are vital to help keep the attrition rate low. Attrition rates can also be high if the stimuli used are difficult to discriminate. Attrition rates for pilot data acquired at the University of Colorado at Boulder for infants with and without hearing loss were much lower than other studies using the CHT paradigm: 10%. Another limitation of infant speech perception assessment methods is that the stimuli used must be rather short in duration, thus precluding the use of sentences or narrative discourse. Performance, even on an “easy” contrast, does not reach 100% for infants. Results of testing must be interpreted with caution – if an infant does not demonstrate discrimination of a contrast, it cannot be assumed that he/she was unable to discriminate it, but perhaps he/she was unable to do the task. Also, results obtained cannot provide information about exactly what the infant perceived, only that a difference was indeed perceived. Therefore, it cannot be inferred that an infant perceives things in the same way as an adult does simply because

he/she demonstrated successful discrimination of a particular contrast. These issues will be discussed further later in this paper.

Although a CHT task such as VRISD is easy to integrate into the audiology clinic and a commercial instrument is available for purchase, the technique lacks thorough research guiding its use in a clinical setting. CHT tasks have been used quite extensively for research, but particular aspects of the test have not been questioned or addressed in published studies (for example, normative data about which speech contrasts to use and for what purposes, and a statistically sound test protocol). In order to develop the CHT/VRISD paradigm to the point where the technique can be incorporated into a clinical protocol for infants and toddlers, this foundational work must first be laid.

Although studies using the conditioned head turn method have yielded interesting information, none have generated normative data or a standardized protocol. VRISD has already been shown to be a technique that can be used successfully with the infant population to assess speech discrimination abilities (e.g., Eilers et al., 1984; Eilers et al., 1977; Nozza, Rossman, Bond, & Miller, 1990). One of the positives of this type of assessment method is the ability to utilize practically any pair of stimuli. Unfortunately, no studies have been completed that demonstrate what "typical" performance is for particular contrasts or provide a comparison of the performance of infants with normal hearing sensitivity to the performance of infants with hearing loss. Nor has any study outlined which contrasts may be

of specific interest for the determination of ability to hear acoustical differences in speech and auditory development. The studies using VRISD for speech discrimination have investigated linguistic questions seeking to study how young children with normal hearing learn language. The studies have rarely focused on perceptual questions to determine a hierarchy of auditory skill development. Most, but not all, studies have used 30 trials in order to determine an infant's discrimination ability. Studies have used differing methods to analyze an infant's performance: percent correct, proportion correct, discrimination index, signal detection theory and criterion measures. The use of different analysis methods makes it difficult, if not impossible, to compare outcomes of different studies. No study has outlined these methods of performance analysis nor has any study compared the methods to one another to determine which method is the best to use with VRISD testing. Because this assessment method holds great potential as a speech discrimination test for clinical use, steps need to be taken to formalize the procedure and reliably quantify and verify performance on the task.

VRISD has been used to investigate many things, including infants' discrimination of voice onset time (Eilers et al., 1981) and vowel duration (Eilers et al., 1984), the effects of stimulus intensity and procedural model on performance measures(Nozza, 1987), and infants' discrimination-in-noise abilities (Nozza et al., 1991; Trehub, Bull, & Schneider, 1981). Although the procedure used for studies utilizing VRISD has primarily been the same, the analysis methods used to assess performance have varied widely. Some

methods reveal performance as a percentage, some as a criterion that must be met, while others use what is termed a “discrimination index”. Different ways of defining performance make meaningful comparison of findings across studies difficult if not impossible.

Before further study can be done to identify the contrasts that should be included in VRISD assessment or what “typical” performance on an specific contrast is, a measure of performance analysis that yields confidence in the results by being based on a strong statistical foundation must be found. Ideally, this measure will also be able to identify an individual infant’s ability to discriminate and decrease the number of trials necessary for testing, making it possible to assess an infant’s ability to discriminate more speech sound contrasts during a test session.

A review and discussion of the various methods that have been used in VRISD testing to analyze performance follows. Comparisons between the assessment methods help to illustrate the difficulty of having differing methods of performance analysis.

## Methods of Performance Analysis

### *Percent Correct*

Percent correct indicates the overall percent correct of an infant’s performance across all VRISD trials. The percent correct method of VRISD results analysis is quite simple to determine:

*Equation 1.  PC = (H + CR) / T*

where PC is percent correct, H is the total number of hits, CR is the total number of correct rejections, and T is the total number of trials.

Kuhl used the percent correct method to analyze results in her study of infants' responses to prototypical and non-prototypical stimuli (Kuhl, 1991). The vowel /i/ was synthesized into many different versions of the vowel by varying formant information. Some of these synthesized vowels were rated as better exemplars of the vowel than others. The author's hypothesis was that infants' perception of speech categories would demonstrate internal structure similar to that of adults. That is, infants would perceive and categorize the vowel in a similar manner as adults, regardless of changes in some of the vowel's formants. The overall percent correct score was obtained for each contrast tested. Then *t* tests were performed to determine if performance differed from chance in each condition tested. Results indicated that infants responded to the speech stimuli similarly to adults, whereas the prototypical (ideal) stimulus served as a sort of "perceptual magnet", resulting in a broader generalization to the perception of the variants of the vowel.

Bohn and Polka (2001) investigated how German adults and infants use spectral and durational cues to perceive German vowels embedded in consonant-vowel-consonant (CVC) syllables. Percent correct was obtained for each contrast. Analysis of Variance (ANOVA) was used to compare performance both within and between subjects on the different contrasts.

Analysis revealed that removing durational cues had a bigger impact on infants' performance than that of adults while infants' use of spectral cues appears to be similar to that of adults. The significant differences found by ANOVA differentiated group performance, but did not provide information about whether children or adults could reach a criterion of mastery for spectral cue discrimination.

Kuhl and colleagues also used percent correct in their study investigating the effects of short term exposure to Mandarin Chinese on English-learning infants' abilities to discriminate a Mandarin speech contrast (Kuhl, Tsao, & Liu, 2003). Infants tested were nine months of age at the onset of the study. Prior research has shown that at approximately this age infants' ability to discriminate speech sounds other than those in their native language decreases significantly (e.g., Werker & Lelonde, 1988). In Kuhl et al.'s study, infants were divided into two groups. One group was exposed to infant-directed Mandarin Chinese in 12 language sessions over four weeks while the other was exposed to infant-directed English in these sessions. Results of VRISD testing indicated that those infants exposed to Mandarin performed significantly better than those infants who were not exposed to Mandarin, thus confirming that even short term exposure to a foreign language could indeed alter speech perception ability. The authors also used signal detection theory to analyze their results. Signal detection theory will be discussed in another section of this paper.

Two aspects of the percent correct scoring method stand out as positive attributes to this scoring method: the scores are easy to calculate and easy to understand.

Unfortunately, the percent correct method has some confounds. If the percent correct method of performance analysis is utilized, the ratio of control to change trials should be 50/50. Many studies use a ratio other than 50/50 in attempt to reduce bias on the part of the tester. In the event that an unequal number of control and change trials occur during testing - if an infant is exposed to 60% control trials and 40% change trials - he/she may score 60% correct even if the infant never completed a head turn during testing. Such a score or result would be quite misleading because the infant would actually not have demonstrated recognition of change for even one trial. Another thing to consider when using percent correct is that the score interpretation is impacted by the task at hand. For example, a score of 90% is more impressive if the task involves four or five choices, rather than two. This is because if there are only two choices, the subject has a 50% chance of guessing correctly whereas if there were five choices, the subject has only a 20% chance of guessing correctly, thereby making a correct response more significant. Also, percent correct indicates not only an infant's ability to perceive the change from background to target stimuli, but also an infant's response bias, which may change over the test session. If, for example, an infant becomes overly interested in the reinforcer, his/her response bias may

become more lenient, thus increasing the number of false positives and result in a lower score.

In Kuhl et al.'s 2003 study, results suggested that there was a significant difference in the performance of the two groups of infants on the Mandarin Chinese speech contrast. Those infants in the group that received exposure to Mandarin scored an average of 65.7% correct while those infants in the English only exposure group scored 56.7%. Although these scores may be significantly different statistically, they may not be of practical significance. In other words, this difference in percentages may not be meaningful. Those infants who were exposed to Mandarin may have scored only one or two more correct than those infants exposed only to English; this significant statistical difference in performance between the two groups may not even be repeatable if another study tried to replicate its results.

*Proportion Correct*

Proportion correct is defined as "the ratio of the correct responses (hits and correct rejections) to the total number of trials" (p. 1929, Nozza, 1987). The calculation of proportion correct is as follows:

$$\textit{Equation 2.} \quad (P)C = [(H / TT) + (CR / TB)] / 2$$

where (P)C is proportion correct, H is the number of correct of hits, TT is the total number of target trials, CR is the number of correct rejections, and TB is the total number of background trials. In CHT testing, scores obtained using

the proportion correct method of analysis are expressed as a number between 0 and 1, with scores closer to one indicating better performance. Proportion correct is analogous to percent correct: a score of 0.5 using the proportion correct method of analysis is essentially equal to 50% using the percent correct method of analysis while a proportion correct score of 0.3 essentially equals a percent correct score of 30%. Because proportion correct is calculated using the proportions of correct responses, the ratio of target to background trials can be something other than 50/50 which helps to reduce tester bias by alleviating the ability to speculate which trial type will be presented next to the subject. A *z* test of means or a *t* test can be used with proportion correct as well as percent correct to determine if performance is significantly better than chance. ANOVA can also be used with both methods to determine if performance on a particular contrast differs significantly from performance on another contrast.

Nozza and colleagues calculated hit rate (correct responses to target trials), false alarm rate (incorrect responses to background trials), and proportion correct for their investigation into infants' ability to discriminate speech sounds in the presence of noise (Nozza et al., 1990). Proportion correct was calculated for individual infants and adults as well the group of infants and the group of adults for each signal-to-noise ratio. Adult subjects were included so that comparisons of performance between adults and infants could be made. The authors found that infants did not perform as well as adults when discriminating speech sound contrasts in the presence of

noise, supporting the notion that infants and young children will be negatively affected by noise more than older children and adults.

Nozza summarized the results from this study as well as others regarding infant speech perception in his paper presented at the A Sound Foundation through Early Amplification conference in 1998 (Nozza, 1998). The purpose of this paper was to explore ways in which hearing loss might affect speech perception. One of Nozza's studies that he discussed in this paper measured speech contrast discriminability at three different intensities: 50, 60, and 70 dB SPL. Performance, measured using proportion correct, was compared to that of adults. It was noted that infants' maximal performance was 0.82, while adults could almost reach 1.0. Nozza points out that the actual proportion correct scores are not the finding of this study to pay particular attention to, but instead attention should be paid to the fact that adults' performance gets better with increased intensity and then asymptotes at only 20 dB HL while infants' performance does not reach an asymptotic level until almost 50 dB HL. These results suggest that in order to reach a level of maximum performance, infants require greater intensity of sound than do adults.

Proportion correct measures have some benefits over other methods of performance analysis. Even numbers of target and background trials are not necessary when calculating proportion correct, unlike percent correct. This allows the ratio of target to background trials to be randomly selected, and is helpful to reduce examiner bias by keeping the examiner unaware of

what type of trial is being played. The score is straight-forward and easy to understand: it ranges from 0 to 1, with 1 being perfect performance.

With proportion correct, like percent correct, interpretation of the score is impacted by the number of choices the task offers. A correct response when there are five choices leaves a subject with only a 20% chance of guessing correctly whereas there is a 50% chance of guessing correctly if there are only two choices. Like the percent correct method, proportion correct indicates not only an infant's ability to perceive the change from background to target stimuli, but also an infant's response bias, which may change over the course of a test session and impact results.

*Discrimination Index*

Discrimination Index (DI) differs from percent correct and proportion correct in that the actual number of opportunities an infant has to demonstrate discrimination is taken into account. The DI is calculated by determining the number of hits (a head turn response to a target trial) minus the number of false alarms (a head turn response to a background trial) divided by the total number of change trials. This equation is:

$$\textit{Equation 3.} \quad DI = (H - FA) / TC$$

where DI is the discrimination index, H is the total number of hits, FA is the total number of false alarms, and TC is the total number of change trials. The

DI is expressed as a number between 0 and 1. The higher the DI score, the better the discrimination ability: perfect performance on the discrimination task results in a DI score of 1. Chance performance is equal to zero (note that this can only happen if there are an equal number of head turns to both the change and the no-change trials). An unequal number of change and no-change trials can be used with DI, which helps to alleviate bias since the testers then do not know the type of trial presented (Eilers & Gavin, 1981).

The DI scoring technique for VRISD was first described by Eilers et al. (1981) in a study investigating whether or not infants could use the onset of voicing to discriminate the voiced and voiceless stop consonants du/tu and ba/pa. A DI was determined for each of 3 blocks of 10 trials. These DI scores were added together and then subjected to a *z* test of means, which determined whether or not performance differed from chance. ANOVA was executed to compare performance on different contrasts. Results failed to find that infants could discriminate a contrast that was cued by only a small difference in the onset of voicing. Infants discriminated naturally produced recorded stimuli, which give the infant more cues, better than synthetically produced recorded stimuli (Eilers et al., 1981).

Discrimination index was also used to evaluate infants' abilities to discriminate differences in the duration of vowels in one, two, and three syllable stimuli (Eilers et al., 1984). Adults participated in the study as well, so that their performance could be compared to that of infants. In any contrast, one stimulus' final vowel was 300 ms in length while the contrast stimulus'

final vowel was 400, 500, or 600 ms long. Percent correct scores were obtained in addition to DI. Both the DI and percent correct scores underwent a *z* test of means to compare performance on a contrast to chance performance as well as ANOVA testing to determine if differences in performance on different contrasts occurred. Results showed that infants could discriminate better as the duration of the vowel increased, indicating that infants could use vowel duration as a cue during speech perception. Although percent correct scores were published, they were not discussed in detail this article. The authors did note that analysis of percent correct scores resulted in the same patterns as those seen in analysis of DI.

Bull and colleagues investigated infants' abilities to detect intensity changes in multisyllabic stimuli (Bull et al., 1984). Infants were presented a multisyllabic speech stimulus contrasted with the same multisyllabic syllable with one syllable slightly louder (2, 4, or 6 dB) than the rest of the stimulus. DI scores were obtained for each infant. A *z* test of means was then performed, as was a *t* test, to assess whether or not performance on a contrast differed significantly from chance. ANOVA was also performed to compare performance on different contrasts. The authors also published results using percent correct to allow the reader to compare the results of the percent correct and DI methods of performance analysis. Results indicated that as the intensity difference between the two multisyllabic stimuli of a contrast increased, DI scores increased, indicating that infants had an easier time detecting larger intensity differences. Results also showed that infants were

capable of detecting differences in intensity that are typically used in speech to signal linguistic stress or suprasegmental cues. The authors did not discuss the percent correct scores and their relation to the DI scores. For the six stimuli for which group results are presented both in DI and percent correct, it appears that DI scores were statistically significant more often than percent correct scores (there are two instances in which results analyzed in percent correct yield a different conclusion than results analyzed in DI).

Bull and colleagues later used multisyllabic stimuli to investigate infants' abilities to detect changes in fundamental frequency between two otherwise identical speech stimuli in an effort to learn more about infants' capacities to detect linguistic stress (Bull, Eilers, & Oller, 1985). DI scores, *z* scores, and ANOVAs were performed to analyze results. Results revealed that infants performed similarly to adults when detecting even small fundamental frequency changes. In conversational speech, these small changes in fundamental frequency may be indicators of linguistic stress. Percent correct scores were also determined, although interpretation of these scores was not included in the article. Further scrutiny of the results of six stimuli for which mean results are provided in both DI and percent correct shows that, on one occasion, results obtained via percent correct were not statistically significant whereas analysis of the results via DI were statistically significant at the $p<0.01$ level.

The discrimination index was used by Oller and colleagues to look into similarities and differences in how infants and adults perceive the shift

between /ba/ and /wa/ along the ba/wa continuum (Oller, Eilers, Burns, & Urbano, 1993). Four contrast pairs of ba/wa were used – each had different durations of the transition from the consonant to the vowel. Previous research by Oller et al. (Oller, Eilers, Miskiel, Burns, & Urbano, 1991) had determined that adults discriminated these pairs with a shorter transition being more discriminable than a longer transition, while the duration of the vowel was inconsequential. The 1993 study set out to replicate this research and follow it with the assessment of infants using the same test procedure for both populations. In addition to calculating the DI for individual subjects, mean DI scores were tested using the *z* test of means to determine if performance was significantly different than chance. ANOVA was also calculated to assess the relative difficulty of the contrasts. Results indicated that sometimes infants' discrimination abilities were poorer than adults' and that patterns of perception of speech sometimes differed as well: infants required approximately three times more difference between the transition durations in the two stimuli being contrasted than did adults.

Discrimination index is an assessment method that allows for an unequal number of change and control trials in testing, which can help to blind experimenter(s) as to the trial type, a method to help alleviate tester bias that has been used in many CHT studies.

Unfortunately, results obtained via DI can be difficult to interpret – a *z* test of means is necessary to interpret the score. Although some studies have reported that both DI and percent correct scores seemingly point to the same

conclusion (e.g., Bull et al., 1984; Bull et al., 1985; Oller et al., 1993), further examination of the data provided shows that the two analysis measures do not always lead to the same conclusion. No studies have gone into detail by outlining the scores for both methods of analysis and comparing the two. Also, the calculation of DI subtracts the number of false positives from correct responses, essentially "penalizing" the infant for false positive behavior. While false positives are not desired, an infant who infrequently false positives is, in essence, demonstrating continued interest in the reinforcer, which is not a bad thing. As in VRA, the occurrence of false positives can be reduced by increasing the length of time between trials and/or by having a well-trained test assistant.

*Criterion*

In the criterion method of performance assessment, testing stops once an infant reaches either a particular number of trials (i.e., 25) or a specific scoring criterion. In most VRISD studies utilizing the criterion method of performance analysis, a floating criterion is used in which discrimination is believed to be present if the subject completes seven out of eight consecutive contrasts correctly. Theoretically, the binomial likelihood of seven correct responses in eight trials is less than 0.05, indicating that the likelihood of this performance occurring by chance alone is less than 5%.

The floating criterion method of analysis was used by Pegg and Werker in their investigation into infants' ability to discriminate a contrast that

differed in phonemes when that difference does not convey meaning in English (Pegg & Werker, 1997). Previous research has shown that young infants can discriminate speech contrasts in both their native language as well as non-native languages but by 10-12 months of age, they perform similarly to adults and can only discriminate speech sounds that are native to their language (Werker & Tees, 1984). Pegg and Werker theorized that if the phonological status of the speech contrast used in this study had a role in discrimination ability, an older group of infants would have difficulty discriminating it while a younger group of infants would not. Two groups of infants were used in this study: one younger (6-8 months) and one older (10-12 months). Twenty-five trials were presented; successful discrimination was determined if an infant reached the criterion during presentation of the 25 trials. The average number of trials it took for infants to reach criterion was not reported. Analysis of Proportions (ANPRO) (an analog of chi-square) was performed to compare the number of infants in each group who reached the criterion of seven out of eight consecutive correct responses. ANPRO analysis revealed a statistically significant difference in the proportion of younger infants who reached criterion than did older infants. Signal detection theory was also used to analyze data from this experiment. A separate section of this paper is devoted to signal detection theory (see page 34). Results of this study indicate that the change in the ability of infants to discriminate non-native speech contrasts seems to be based on the

phonological status of a contrast in their native language, that is, whether or not it has meaning in their native language.

Anderson and colleagues also used the criterion method to analyze performance in a CHT task (Anderson, Morgan, & White, 2003). The authors hypothesized that the manner in which phonetic categories in an infant's native language emerge would be similar to the regularity with which they appear in the native language. Therefore, contrasts from non-native languages that relate to categories of sounds that occur more frequently in the native language would disappear earlier. The investigators used Hindi and Salish sound contrasts to assess this hypothesis: due to the higher frequency of coronals (those phonemes that are made by putting the tongue on or near the upper teeth or the alveolar ridge of the mouth, e.g., /t/, /d/) in English, discrimination of non-native coronal contrasts should decline earlier than discrimination of non-native dorsal contrasts (those phonemes that are made by putting the body of the tongue on or near the roof of the mouth or soft palate, e.g. /g/, /h/).

The authors conducted a series of experiments to investigate their hypothesis. The first tested eight and nine month olds' abilities to discriminate a Hindi contrast (a coronal retroflex/dental stop contrast) and a Salish contrast (a dorsal velar/uvular ejective contrast). Results of this study suggested that infants whose native language is English can discriminate the non-native dorsal contrast better than the non-native coronal contrast, consistent with the investigators' hypothesis that non-native contrasts which

correspond to a less frequently occurring English phonological category are more discriminable to infants than those non-native contrasts which correspond to a more common English phonological category.

The second experiment tested six and seven month olds' abilities to discriminate the same contrasts outlined in the first experiment. The authors observed that these younger infants had a slower head turn response than did the older infants. These younger infants discriminated the two contrasts similarly well. This supports previous work by Werker and Tees (1984) indicating that infants' ability to discriminate non-native contrasts changes during the first year of life and is also consistent with the authors' hypothesis that the coronal category is acquired earlier, presumably due to its more frequent occurrence in the English language; consequently, a decline in discrimination ability appears earlier for the non-native coronal contrast than for the non-native dorsal contrast. In both experiments, most infants who achieved the seven-out-of-eight criterion did so in 15 trials or fewer.

The criterion method of performance analysis is essentially "adaptable" to each infant – test time is shorter if the infant is good at discriminating a contrast. This shorter test time for each successfully discriminated contrast allows more contrasts to be assessed in one session than has traditionally been able to be tested. As noted by Anderson et al. (p. 166, Anderson et al., 2003) and commonly observed by pediatric audiologists, the more time that is needed for testing, the less likely the infant is to finish the test.

Before a criterion method such as seven correct out of eight consecutive trials can be adopted for performance analysis, some things need to be taken into consideration. For instance, an infant could perform at a level higher than the criterion, yet fail to meet it (e.g., an infant gets six in a row correct, then misses two, then gets six more correct when the criterion is seven out of eight in a row). When a floating criterion is used, the *p*-value increases according to the total number of trials. In other words, while the binomial likelihood of reaching seven out of eight consecutive correct responses in nine trials is .049, in 10 trials it is .061 (and no longer statistically significant) (Anderson et al., 2003). Therefore, the criterion of seven out of eight consecutive correct responses means something different if it occurs in the context of eight trials versus in the context of 18 trials. Also, Anderson et.al. (2003) report that a floating criterion may not be appropriate for scoring individual subjects due to the fact that the likelihood of achieving seven out of eight correctly by chance in, for example, 25 trials is .23 (that is, this performance would occur by chance 23% of the time).

*Signal Detection Theory*

Signal Detection Theory (SDT) is a statistic typically referred to as an index of sensitivity to a signal. Underlying this model are two normal distributions, one representing the signal, the other representing the noise (in general terms, this noise is simply variability or uncertainty). The result of the calculation of SDT is typically referred to as *d'*, which represents how well one

can discriminate between the two distributions. *d'*, the standardized difference between the means of the two distributions, is calculated by:

*Equation 4.* $d' = z(FA) - z(H)$

where z(FA) is the z score of the total number of false alarms and z(H) is the z score of the total number of hits. Therefore, in order to calculate *d'*, both the hit rate and false alarm rate of the subject need to be known. A *d'* of zero indicates that the subject could not distinguish the signal (in the case of VRISD, the target) from the noise (the background). As *d'* increases, the ability of the subject to detect the signal from the noise increases. SDT takes into account the subject's response bias (β) (a subject's likelihood of saying 'yes' or 'no'). The hit rate and false alarm rate also help to determine the response bias. Figure 1 represents this sensitivity index.

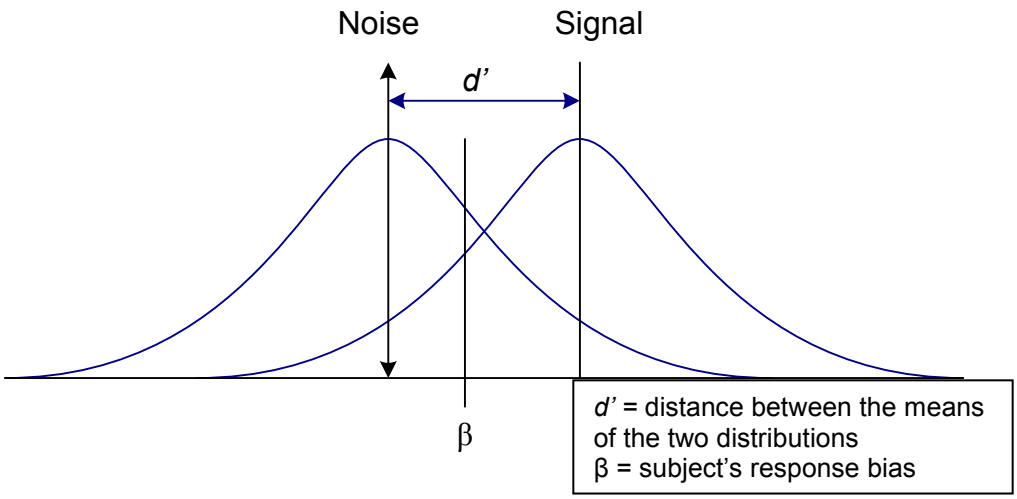


Figure 1. Depiction of signal detection theory sensitivity index.

The two distributions may or may not overlap: the degree to which the two overlap depends on how sensitive a subject is to the signal. The more sensitive the subject is to the signal, the further away from each other the two distributions will be (and the larger the value of *d'*). On the other hand, if a subject has great difficulty telling the target signal from the noise, the two distributions may completely overlap (and the smaller *d'* will be, perhaps as small as zero). A negative *d'* is possible, but rare. A negative *d'* suggests that the subject gave frequent false positives (that is, the subject's bias was very low, so he/she responded frequently).

Kuhl et al.'s study of infants exposed to Mandarin Chinese (discussed earlier) calculated *d'* in addition to percent correct to analyze results (Kuhl et al., 2003). The published results of Kuhl et al.'s study do not give actual *d'* numbers, it only states that results obtained using this method of analysis yielded the same pattern of results as those obtained using percent correct.

A' is typically referred to as a nonparametric statistic similar to *d*' (Wickens, 2002). The response bias of subjects is controlled when using A' because it is a measure that considers the number of false alarms relative to the number of hits, as does *d*', but uses only one pair of hit and false alarm values. Because only one pair of data is available for analysis, Wickens points out that "considerable extrapolation" is necessary (p. 71, Wickens, 2002). The value of A' ranges from 0.5 (performance is equal to chance) to 1.0 (performance is perfect). Should an A' of less than 0.5 result, it is assumed that a strategy other than that needed for the task was used. The

formula to calculate A' when the number of hits is greater than or equal to false alarms is

$$\textit{Equation 5.} \quad A' = 0.5 + (H - FA)(1 + H - FA) / [4H(1 - FA)]$$

where H is the proportion of hits and FA is the proportion of false alarms (Macmillan & Creelman, 2005). When the number of hits is less than or equal to the number of false alarms, the equation used to calculate A' changes to

$$\textit{Equation 6.} \quad A' = 0.5 - (FA - H)(1 + FA - H) / [4FA(1 - H)]$$

Pegg and Werker (1997) used A' in addition to the criterion method to analyze their study's data. The use of signal detection theory for analysis of performance is useful in that comparison of different studies can be made, regardless of the method of testing or number of trials used (Green & Swets, 1966). Signal detection theory results are not impacted by the design of the task being measured nor are they impacted by response bias.

Unfortunately, signal detection theory is not familiar to many people. A *d'* (or A') score itself is not easily interpreted. It also requires the use of many trials in order to determine the statistic accurately (Wickens, 2002), which is difficult if not impossible with infants. It also requires the use of underlying statistical assumptions regarding the population and the distribution of

responses, particularly that the underlying distribution is assumed to be Gaussian.

Although A' has been referred to as a nonparametric statistic based on *d'*, authors point out that in actuality, A' is indeed parametric, relying on the same assumptions of the underlying distributions that *d'* does (Pastore, Cawley, Berens, & Skelly, 2003; Wickens, 2002). The authors also indicate that A' can sometimes underestimate sensitivity of a subject. A', like *d'*, is not easily interpreted by most people.

## Comparison of Analysis Methods

To assist in understanding how these different performance analysis measures can lead to different conclusions even when based on the same data, data obtained in a pilot study done at the University of Colorado at Boulder was analyzed.

Pilot data was obtained from both normal hearing infants and infants with hearing loss at the University of Colorado at Boulder's Center for Speech, Language and Hearing. Software, similar to the original software used by Eilers et al. (1977) was developed for VRISD testing by William J. Gavin, Ph.D., who also developed the original software.

In the pilot study, infants were tested on the contrasts /a/-/i/, /a/-/u/, /u/-/i/, and /s/-/sh/. Each contrast tested consisted of 30 trials and was presented at 60dB SPL. The data obtained from these 30 trials was then analyzed post hoc in each of the performance analysis methods described above. Table 1

outlines results for three subjects analyzed in each of the various performance analysis methods outlined above.

| | Subject A | Subject B | Subject B-2 | Subject C | Subject C-2 | Subject C-3 |
|---|---|---|---|---|---|---|
| Contrast assessed | a/i | a/i | u/a | sa/sha | i/u | a/i |
| Percent correct | 80%* | 83%* | 86%* | 42% | 55% | 75%* |
| Proportion correct | 0.80* | 0.83* | 0.86* | 0.42 | 0.55 | 0.75* |
| Discrimination Index | 0.6* | 0.5* | 0.73* | 0 | 0.2* | 0.66* |
| *d'* | 2.15* | 2.76* | 2.08* | -0.37 | 0.51 | 2.2* |
| Criterion (7 out of 8) | Finished after trial 14* | Finished after trial 9* | Finished after trial 8* | Finished all 30 trials | Finished after trial 9* | Finished after trial 14* |

* = score indicates successful discrimination of contrast assessed

Table 1. Performance analysis results for three subjects.

The three subjects are labeled Subject A, Subject B, and Subject C. Subjects were 9 to 12 months old. Subjects A and B (males) had normal hearing sensitivity while Subject C (female) had a moderate–severe sensorineural hearing loss and was fit with hearing aids at 6 weeks of age. This subject wore her hearing aids during VRISD testing. All subjects' families spoke English in the home and no other disabilities were present.

Those results presented in Table 1 that indicate successful discrimination of the contrast are marked by an asterisk. To determine if

results obtained in percent correct, proportion correct, and discrimination index are significantly different than chance, the scores are subjected to a z test of means. The criterion method of analysis used a criterion of seven out of eight in a row correct as a demonstration of ability to discriminate, as has been used in most published VRISD studies that use a criterion (e.g., Anderson et al., 2003). Beginning with trial 1, subject responses were determined to be either 'correct' or 'incorrect'. Once the criterion of 7 trials out of 8 in a row correct was met, testing could stop. If the criterion was not met, the subject would complete all 30 trials during testing.

To compare the different performance analysis measures for the results, details regarding the contrast tested or the subject are not necessary. What is important is to study the differing performance analysis measures and what each says about the performance of the subject and how these different measures compare to one another.

According to the percent correct, proportion correct, DI, and SDT methods of analysis, only subjects A, B, B-2, and C-3 successfully discriminated the speech contrast assessed. Both the discrimination index and the criterion method of analysis demonstrate that subject C-2 also performed well for the contrast tested, whereas analysis using percent correct indicates otherwise. All analysis methods are in agreement that subject C did not discriminate the contrast successfully.

The results obtained and scored in *d'* are more difficult to interpret. Typically, *d'* scores range from 0 to approximately 4.0. The larger *d'* is, the

better the subject's performance on the task. The signal detection theory statistics are useful for comparing performances between contrasts and/or subjects, but aren't very informative/intuitive when analyzing the performance of an individual subject.

It is of particular interest to note subject C-2, where the different analysis methods result in differing interpretations from one another. For example, subject C-2 would have completed testing after trial nine had the criterion method of performance analysis been used. But when performance of all 30 trials of testing was analyzed (as is the case for percent correct, proportion correct, and discrimination index), the performance scores decrease. This decrease in performance may have occurred due to subject fatigue or boredom with testing, which was noticed somewhat frequently in our pilot data test sessions, especially in the latter half of assessment of a contrast. In this case, the method of using a criterion to assess performance appears to give a better depiction of the subject's ability.

It is also interesting to study subjects A, B, and C-3. In these cases, the subjects were assessed on the same contrast (/a/ vs. /i/) and seemingly performed very similarly to one another in all the performance analysis methods.

Statistical analysis of VRISD performance is necessary in order to define when an infant's response reaches a level that could be defined as a 'pass' (i.e., the infant can discriminate the contrast). As reviewed above, current methods of performance analysis allow for statistical analysis of the

performance after testing. For example, a *z* test can be used with most methods of performance analysis to answer the question: Is performance better than chance? ANOVA can be used as well, which allows for comparison of performance on contrasts to determine if performance on one contrast is better or worse than performance on another contrast. Unfortunately, with so many different definitions used for measuring performance in VRISD testing, these statistical techniques (except for SDT) cannot be applied across studies, unless the studies have been replicated (in terms of measurement of performance).

Other aspects of testing that must be considered before deciding upon a method of performance analysis for VRISD are: reliability of the performance measure and the attention span and needs of infants. A test such as VRISD must have a method of performance analysis that is reliable if it is to be used as a measure of an infant's perceptual abilities. The use of a valid statistically-based technique will help to achieve this reliability. The performance measure must also be able to be utilized with different populations – for example, with infants and toddlers with normal hearing as well as those with hearing loss. The length of time for which an infant will cooperate for testing will depend upon many factors such as the time of day, whether he/she is hungry, whether he/she has a clean diaper, et cetera. Because infants have limited ability to sit and complete a task such as VRISD, using test time efficiently by maximizing the amount of information obtained during the test session is critical.

Many studies that use the VRISD paradigm use 30 trials to determine an infant's perceptual abilities. One limitation of a visual reinforcement task is that infants may become habituated to the reinforcer before testing is completed (Schmida, Peterson, & Tharpe, 2003). Schmida et al. counted the number of head turn responses until habituation in VRA (2003). The average number of head turn obtained from the group of infants in their study was 15. Other studies have also counted the number of head turns obtained during VRA testing, which ranged from an average of 11 (Thompson, Thompson, & Vethivelu, 1989) to an average of 16 head turns (Culpepper & Thompson, 1994). In VRISD pilot studies using 30 trials, we have noted that there are a fair number of infants who do quite well during the first half of testing, only to become seemingly uninterested in testing during the latter half of the test session. Due to the limited number of head turns that can be obtained prior to many infants becoming habituated to the task, it is important to maximize the information obtained during testing while considering the length of the test session. Unfortunately, in situations where the infant becomes bored or habituated to the VRISD task, the use of most of the performance analysis methods described in the above sections would suggest that the infant performed poorly even though he/she may have scored seven correct out of the first eight trials or subjects may be dropped from the study because they became disinterested or began to cry, resulting in an increased attrition rate. The frequency of this type of occurrence in our pilot studies has reinforced our belief for the need of a VRISD protocol that not only has a reliable statistic

used as its measure of performance, but also for a protocol that can be shortened, if perhaps only for those infants who perform above a certain level for the first *X* number of trials (as did subject C-2).

Typically, VRISD testing is composed of background and target trials as well as background sounds that are not actually part of a background trial. The background stimulus repeats throughout testing except for when a trial is initiated by the examiner. A trial is initiated once the subject is centered and determined to be ready for testing. The trial type is determined by the computer semi-randomly. The trial may be either a background trial or a target trial. A correct response for a background trial is no head turn, while a correct response on a target trial is a head turn. Once the trial is complete (typically a trial consists of the stimuli repeating three times), the background stimuli returns. This background stimulus is played until the examiner again determines that the subject is ready for a trial. These extraneous background stimuli serve no purpose other than to allow time for the subject to return to center. One way to decrease test time would be to rid the current VRISD paradigm of the extraneous background stimuli that are not scored as part of the VRISD test itself. Removing these extra stimuli could potentially reduce test time by half, but could also introduce the possibility for error should the subject be rewarded with a correct response (either a hit or a correct rejection) even when they are not attentive to the task (for example, the infant may be reaching for an interesting toy, talking, or crying).

Researchers at House Ear Institute have been developing a battery of tests to assess speech perception in infants and toddlers (Eisenberg, Martinez, & Boothroyd, 2007). This battery includes a version of VRISD (termed VRASPAC – Visual Reinforcement Audiometry Speech Pattern Contrast) that both removes the extraneous background stimuli and uses a statistically-based formula for determining performance (Boothroyd, 2004; Eisenberg et al., 2007; Martinez, Eisenberg, Boothroyd, & Visser-Dumont, 2008). The VRASPAC has been developed along with several other tests of auditory perceptual abilities, each of which has been developed for particular age groups. VRASPAC is for use with infants and toddlers and utilizes vowel-consonant-vowel (VCV) stimuli. It has been used on both normal hearing and hearing impaired infants and toddlers with some success. Though the VRASPAC is promising, its statistical basis for determining adequate performance on a contrast has not yet been thoroughly provided or explained, nor has it yet been shown to be a valid, reliable test for the assessment of infant speech perception. Recently, Uhler used the VRASPAC for her longitudinal study of speech perception abilities of young cochlear implant recipients (Uhler, 2008b). She noted that there were instances that were counted as correct rejections by the software, even though the subject was not centered, was fussy, or talking, but was not looking at the reinforcer (Uhler, 2008a). This situation could be avoided by having every trial, both targets and backgrounds, initiated by the tester.

The CHT paradigm for assessing infants' abilities to perceive differences in sound has proven itself to be a tool with great potential for clinical use with infants. The foundational work for this paradigm must be further explored and validated before it can become a clinical tool. Once a reliable method of performance analysis is realized, the questions of which phoneme contrasts to use and when to use them as well as how many trials are necessary to assess performance can be addressed. With a method of performance analysis that allows for not only analysis of performance on an individual contrast for an individual infant, but also a comparison of performance on differing contrasts and comparison to others' performance, VRISD has the potential to reveal information about an infant's speech perception abilities. This information would add to the fund of knowledge regarding speech perception skills and abilities of infants and toddlers with and without hearing loss as well as be applied to habilitation goals, validate benefit of amplification, and help determine cochlear implant candidacy for infants with hearing loss.

The current study was designed to investigate performance of infants on a speech discrimination task using the VRISD procedure. This paper aims to address two questions:  How do normal-hearing infants perform on three contrasts? What is the most appropriate statistic for determining a stopping criterion?

# CHAPTER 3 -- METHODS

## SIMULATED DATA FROM PSEUDOSUBJECTS

Initially, simulated data for pseudo subjects were created so that the statistical methods could be explored prior to the acquisition of data on infants. Fifteen pseudo subjects were created whose VRISD performance ranged from correct responses on every trial to incorrect responses on every trial. One pseudo subject performed perfectly (getting all 30 trials correct). One pseudo subject performed incorrectly on all 30 trials. Performance for the remaining 13 pseudo subjects was modeled after trends seen in results obtained during pilot data acquisition (see Appendix P for individual pseudo subject response patterns). As discussed earlier in this paper, many methods of performance analysis have been used with the VRISD paradigm. The methods presented earlier (percent correct, proportion correct, signal detection theory, criterion and discrimination index) were used to assess performance of the pseudo subject data. An additional method of analysis, binomial probability, was also used. What follows is a brief description of each of these methods.

Percent correct – Percent correct was determined by dividing the number of trials which were responded to correctly (that is, hits and correct rejections) and dividing that number by the total number of trials. Technically, this method should be only be used when an equal number of target and background trials are presented. In terms of interpreting VRISD performance,

the percent correct score reveals the percentage of trials on which a subject performed correctly. Scores range from 0–100%.

Proportion correct – When determining proportion correct, the proportion of correct responses to target trials is added to the proportion of correct responses to background trials, and is then divided by two. The proportion correct score is analogous to the percent correct score: it reveals the proportion of trials on which a subject performed correctly. The proportion correct score ranges from 0.0 to 1.0.

Signal detection theory – d' is a measure of signal detection theory. Scores typically range from 0 to infinity. Rarely, a negative score will arise. This occurs when the subject has a bias towards responding, even when there was no stimulus change. As the SDT score gets bigger, performance gets better. d' has frequently been used to compare VRISD performance between subjects, groups, and/or contrasts.

Discrimination index – The discrimination index score is determined by subtracting the number of false positives from the number of hits and then dividing by the total number of target trials. This score is a number typically between 0 and 1, but may be a negative number if a child responds with many false positives. This DI score is then subjected to a z test of means in order to determine if performance is significantly better than chance (which has a DI score of 0.0). A z score of greater than 1.96 indicates that performance was indeed better than chance and that the subject has successfully discriminated the contrast being assessed.

Criterion – The criterion method that has been most used in published VRISD studies is that of seven correct responses out of eight consecutive trials (e.g., Anderson et al., 2003). Once this criterion has been met, testing is terminated. Should this criterion of performance not be met, testing ends at some predetermined number of trials (for example, 25). Meeting the criterion results in the assumption that the subject can indeed discriminate the contrast being assessed.

Binomial Probability – This method of analysis of performance has not been used in any published VRISD studies. The probability of an occurrence can be termed a binomial probability if the following hold true: there are two possible outcomes for each trial, the probability of each outcome remains constant from trial to trial, each trial is independent of one another, and the two outcomes are complementary to one another. That is, the probability of the outcomes for each trial, *p* and *q,* add up to one. Eilers and Gavin demonstrated that the VRISD task can indeed be approached as binomial (1981). In VRISD, there are four possible outcomes as it is a two-alternative task: two correct and two incorrect outcomes. One type of correct outcome is a 'hit', where the infant correctly turns his/her head to a change (target) trial. Another correct outcome is a 'true negative', where the infant correctly does not turn his/her head to a no-change (background) trial. One type of incorrect outcome is a 'miss', where the infant does not turn his/her head to a change trial. The other type of incorrect outcome is a 'false positive', when the infant turns his/her head during a no-change trial. These four outcomes can be

visualized as a 2 x 2 contingency table (also referred to as a confusion matrix) which is depicted in Table 2:

| | Change Trial | No-Change Trial |
|---|---|---|
| Head Turn | Hit | False Positive |
| No Head Turn | Miss | True Negative |

Table 2. 2 x 2 contingency table for VRISD.

The probability of a correct response occurring on any particular VRISD trial is 0.5. Binomial probability can be assessed sequentially, as a probability on successive trials, to determine the probability of achieving success on consecutive trials. While the probability of getting the first trial correct is 0.5, the probability of getting both the first and the second trials correct is 0.25 and the probability of getting the first, second, and third trials correct is 0.125. The likelihood of performing well on consecutive trials decreases as the number of trials increases. The formula for determining this cumulative binomial probability is:

*Equation 7.* $$F(a) = \sum_{x \leq a} P(X = x)$$

In addition to using binomial probability of performance based on all 30 trials, a criterion based on the binomial probability cumulative distribution

function (cdf) was also used to analyze performance with the goal of decreased test time by decreasing the number of trials necessary when assessing infant speech perception as well as to ensure statistical significance (that is, the infant's performance could not have occurred by chance). Meeting this criterion would indicate that the subject successfully discriminated the two speech sounds from one another. The criterion was set at two consecutive trials that had a binomial probability of success with a significance value of $p<0.05$, with a minimum of eight trials, two of which must be target trials. Eight trials was chosen as the minimum number of trials to be completed to ensure subjects were exposed to both control and target trials, even if the unlikely scenario of three consecutive background trials followed by one target trial followed by three consecutive background trials were to occur (recall that the software limits the number of consecutive background trials to three). In order to be able to cease testing after the eighth trial, the significance values of both Trial 7 and Trial 8 must have been less than or equal to 0.05, and responses to the two target stimuli must have been correct. Given these restrictions, an infant could respond incorrectly to two of the 8 trials and maintain this significance value. The requirement of the significance values of the last two trials to be less than or equal to 0.05 ensures that performance was significantly better than chance regardless of the number of trials presented. Should this criterion method work well for VRISD assessment, it would decrease the number of trials necessary to present when testing infants. This decrease in the number of trials would be achieved

while maintaining statistical significance, leaving the tester confident that the infant could indeed discriminate the contrast on which he/she was tested. This would ensure an accurate conclusion regarding the ability of infants who may otherwise stop responding due to fatigued or boredom during the traditional presentation of 30 trials. Test time would be decreased and more speech contrasts could be assessed in one test session.

Performance for each contrast was scored in each of the above methods, which then were compared to one another. With any of these scoring methods, performance that exceeded that expected by chance indicated that the subject could indeed discriminate the speech contrast being assessed.

Since pseudo subject performance ranged from perfect to completely wrong, no real 'trend' in performance between subjects could be seen. This is not what we have experienced in our pilot studies. In both published studies and our pilot studies, trends in performance on any particular contrast can be seen.

Pseudo subject 7 is an example of an infant who stays on task and demonstrates discrimination throughout testing. Figure 2 shows Pseudo subject 7's performance as measured in percent correct and binomial probability over 30 trials. Both result in the conclusion that this subject can successfully discriminate the contrast being assessed.

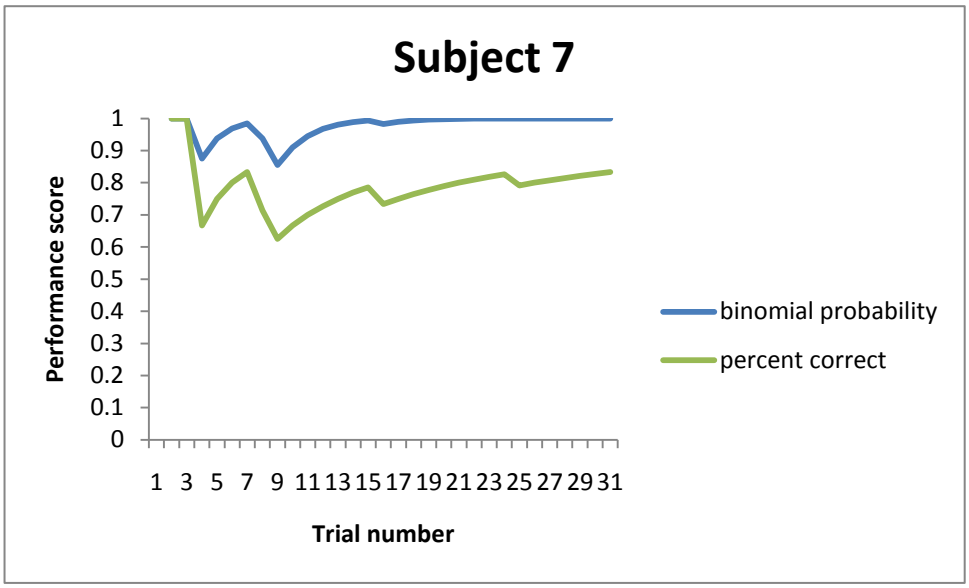


Figure 2. Pseudo subject 7's performance measured in percent correct and binomial probability over 30 trials.

Table 3 depicts Pseudo subject 7's performance in each of the performance analysis measures. An asterisk beside a score indicates successful discrimination based on the requirements of that particular performance measure. Each of the performance analysis measures indicates that the subject can discriminate the contrast.

| | |
|---|---|
| Binomial Probability | P<0.01* |
| Percent Correct | 83%* |
| Proportion Correct | 0.83* |
| Discrimination Index | 0.67 (z=3.9)* |
| Criterion: 7/8 | Done after trial 16* |
| Criterion: Binomial Probability | Done after trial 12 (p<0.05*) |

Table 3. Pseudo subject 7's performance according to each of the different performance analysis measures.

While sometimes the performance analysis methods agreed as to the performance outcome for a contrast, there were also situations where they

failed to adequately describe performance for some subjects, primarily those whose performance diminishes during the second half of the test session. This diminishing performance was seen in CU Boulder pilot studies as infants became bored or disinterested in the reinforcer and the test itself. For example, Figure 3 depicts decreasing performance over 30 trials:

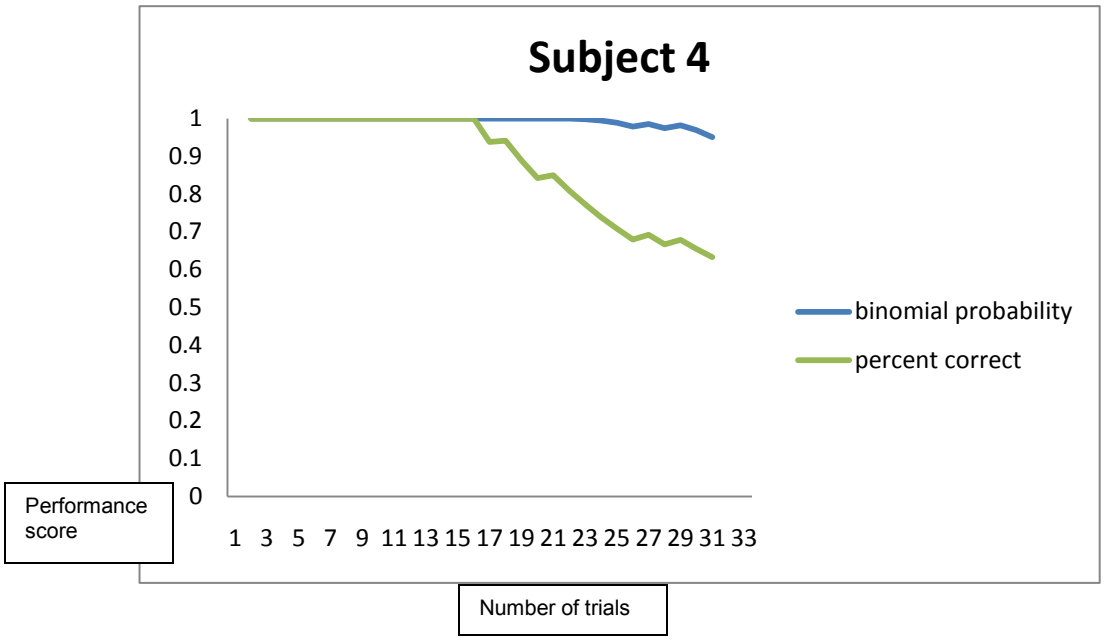


Figure 3. Pseudo subject 4's performance measured in percent correct and binomial probability over 30 trials.

After 30 trials, analysis by percent correct results in the decision that the subject did not successfully discriminate the contrast whereas the probability that the subject could have performed as well as he/she did by chance is below 5%. Table 4 displays this subject's results in the various analysis methods:

| | |
|---|---|
| Binomial Probability | P<0.05* |
| Percent Correct | 63% |
| Proportion Correct | 0.63 |
| Discrimination Index | 0.31 (z=1.8) |
| Criterion: 7/8 | Done after trial 8* |
| Criterion: Binomial Probability | Done after trial 8 (p<0.01*) |

Table 4. Pseudo subject 4's performance according to each of the different performance analysis measures.

Again, an asterisk beside a score indicates successful discrimination based on the requirements of that particular performance measure. A different conclusion is drawn about this subject's performance depending on the method of scoring used. Percent correct, proportion correct, and DI scores result in the conclusion that the subject did not successfully discriminate the contrast whereas probability and the two criterion methods conclude that the subject did indeed discriminate the contrast. The two criterion methods decrease the number of trials needed by two-thirds, thereby significantly reducing test time.

Pseudo subject 15 is another example of a subject whose performance diminishes as testing continues (Figure 4).

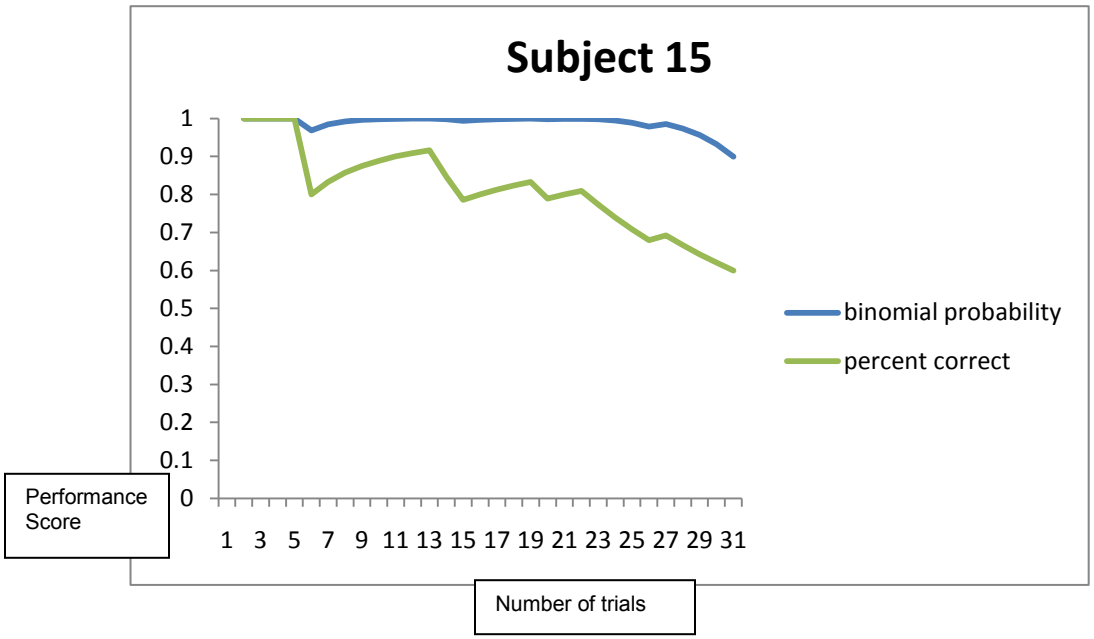


Figure 4. Pseudo subject 15's performance measured in percent correct and binomial probability over 30 trials.

Here, neither the binomial probability nor percent correct indicate successful discrimination of the contrast. Discrimination index and proportion correct do not point to successful discrimination either (see Table 5). But if we look at the two criterion measures, the subject would have finished testing after no more than 8 trials. Again, test time would be significantly reduced if the criterion measures were used to measure performance.

| | |
|---|---|
| Binomial Probability | P<0.10 |
| Percent correct | 60% |
| Proportion Correct | 0.58 |
| Discrimination Index | 0.25 (z=1.47) |
| Criterion: 7/8 | Done after trial 8* |
| Criterion: binomial probability | Done after trial 8 (p<0.02*) |

Table 5. Pseudo subject 15's performance according to each of the different performance analysis measures.

Pseudo subject 5 is an example of a subject who is an intermittent performer, sometimes staying on task, sometimes not:

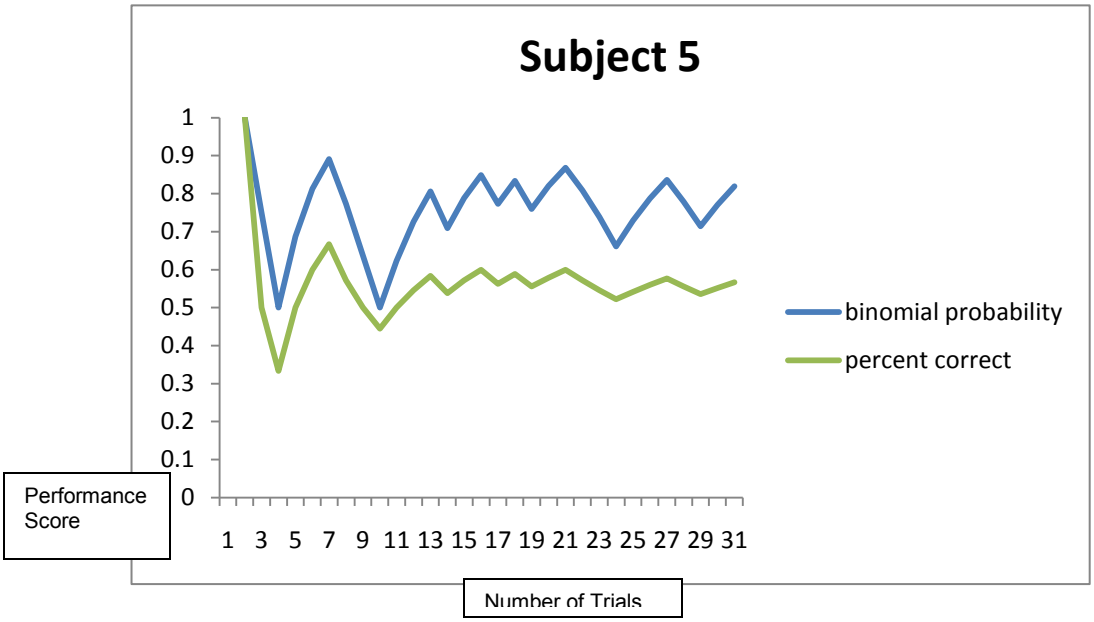


Figure 5. Pseudo subject 5's performance measured in percent correct and binomial probability over 30 trials.

As seen in Table 6, all methods of analysis agree that this subject did not perform better than chance. Even in the criterion methods of analysis, the subject would have completed the entire test of 30 trials.

| | |
|---|---|
| Binomial Probability | P=0.19 |
| Percent correct | 57% |
| Proportion correct | 0.56 |
| Discrimination Index | 0.19 (z=1.1) |
| Criterion: 7/8 | Complete all 30 |
| Criterion: Binomial Probability | Complete all 30 |

Table 6. Pseudo subject 5's performance according to each of the different performance analysis measures.

Pseudo subject 12 is another example of a subject for whom all methods of analysis agree that performance was not better than chance.

Figure 6 depicts this subject's performance as measured in percent correct and in binomial probability over 30 trials. Table 7 shows pseudo subject 12's performance as measured by the different measures of performance analysis. All measures agree that this subject did not discriminate this contrast. The subject would have completed all 30 trials for the two criterion methods, as performance never met the defined criterion.

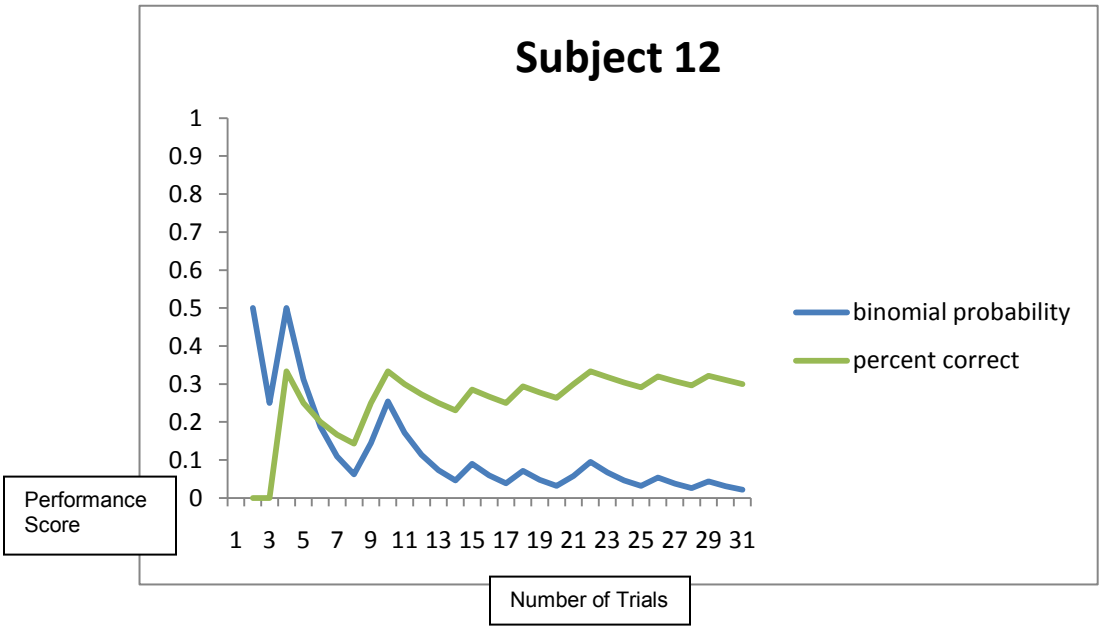


Figure 6. Pseudo subject 12's performance measured in percent correct and binomial probability over 30 trials.

| | |
|---|---|
| Binomial probability | P=0.98 |
| Percent correct | 30% |
| Proportion correct | 0.29 |
| Discrimination Index | -0.17(z=0.98) |
| Criterion: 7/8 | Complete all trials |
| Criterion: binomial probability | Complete all trials |

Table 7. Pseudo subject 12's performance according to each of the different performance analysis measures.

Some VRISD subjects start the test session performing inconsistently, but their performance improves after the first few trials. Pseudo subjects 6 and 9 illustrate this trend.

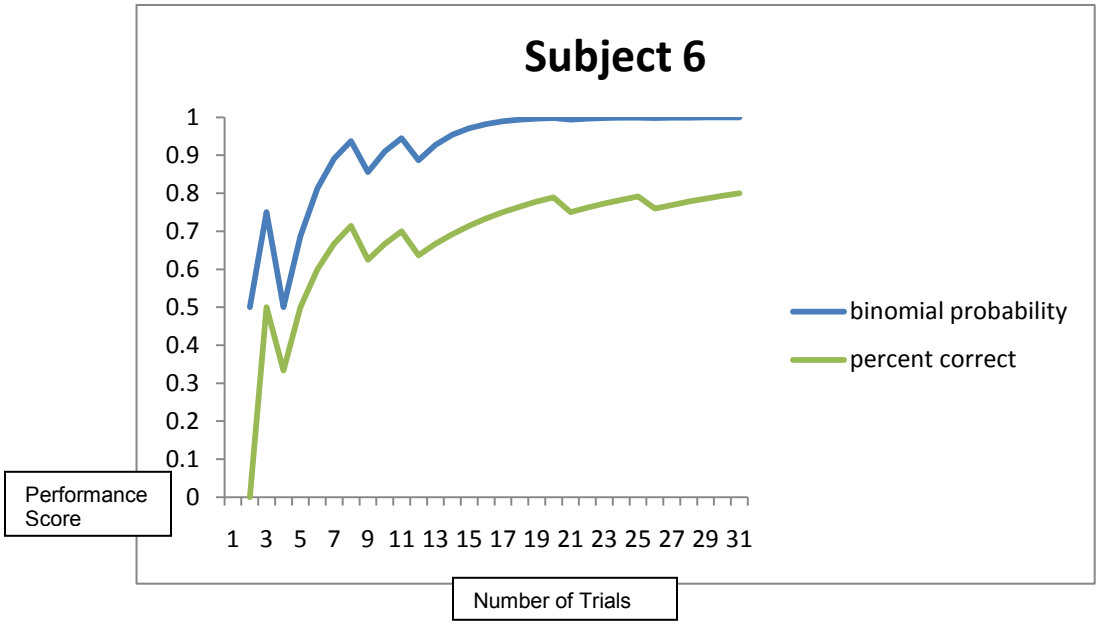


Figure 7. Pseudo subject 6's performance measured in percent correct and binomial probability over 30 trials.

| | |
|---|---|
| Binomial Probability | P<0.01* |
| Percent correct | 80%* |
| Proportion correct | 0.8* |
| Discrimination Index | 0.6 (z=3.5*) |
| Criterion: 7/8 | Done after trial 17* |
| Criterion: Binomial probability | Done after trial 14 (p<0.03*) |

Table 8. Pseudo subject 6's performance according to each of the different performance analysis measures.

Figure 7 depicts pseudo subject 6's performance as measured in percent correct and binomial probability over 30 trials. Table 8 reveals results of pseudo subject 6's performance as measured by the different

performance analysis measured. All methods of performance analysis agree that pseudo subject 6 has demonstrated successful discrimination of the contrast, although analysis by the criterion methods would result in decreased test time.

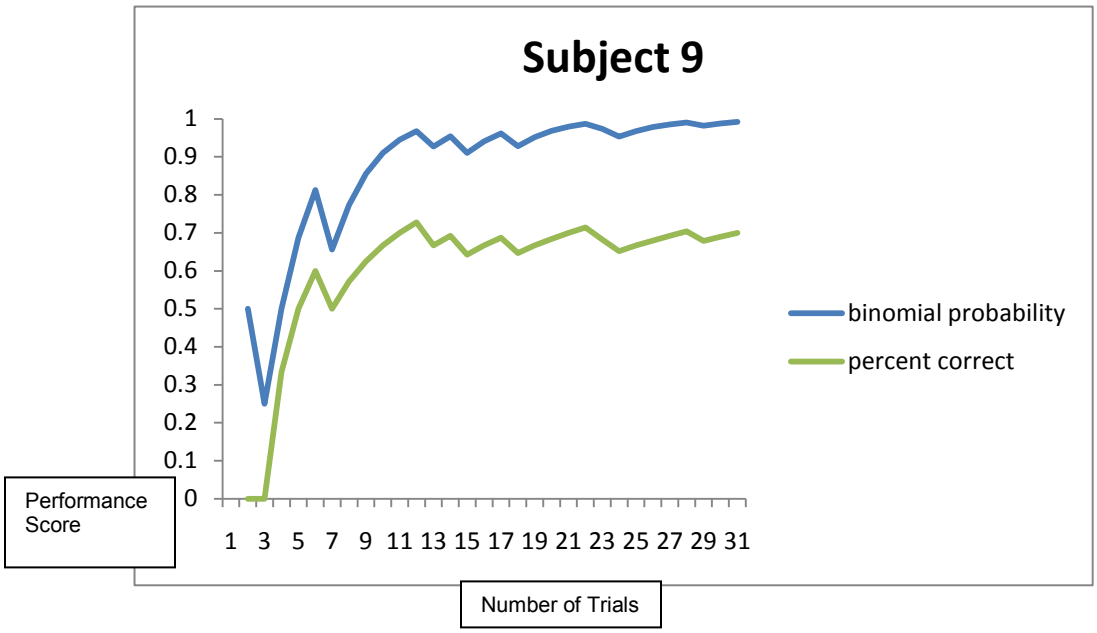


Figure 8. Pseudo subject 9's performance measured in percent correct and binomial probability over 30 trials.

| | |
|---|---|
| Binomial Probability | P<0.01* |
| Percent correct | 70%* |
| Proportion Correct | 0.70* |
| Discrimination Index | 0.44 (z=2.57*) |
| Criterion: 7/8 | Done after trial 10* |
| Criterion: Binomial Probability | Done after trial 19* |

Table 9. Pseudo subject 9's performance according to each of the different performance analysis measures.

Pseudo subject 9 also demonstrates the trend of improving performance as testing continues (Figure 8). All methods of performance

analysis reach the same conclusion: successful performance on the task (Table 9). Again, the use of a criterion measure would result in decreased test time.

### Conclusions from Simulated Data

The analysis of simulated data allowed for exploration of the various methods of performance analysis that have been used with VRISD as well as the development of a new method: a binomial probability based criterion. This probability-based criterion appears to have potential for use with VRISD assessment. These analyses were compared to data attained from 15 infant subjects.

## PARTICIPANTS

The number of participants required for this study was obtained through power analysis. Choosing a medium effect size, power in a study using 15 participants is 0.63. One way to interpret this figure is that there is a 37% chance of making Type II error (failing to reject the null hypothesis that performance is no different than chance when indeed the null hypothesis is incorrect) with 15 participants. While increasing the number of participants to 20 decreases the likelihood of making a Type II error to 29%, we feel that being fairly conservative and therefore mistakenly failing to acknowledge a difference in performance that is actually present is preferable to finding a difference in performance when the child actually cannot detect the difference

in the sounds. Also, this study is essentially an exploration of performance analysis methods. Insight gained from this study will be addressed in greater depth in future studies.

Fifteen infants (6 males, 9 females) with normal hearing sensitivity were tested for this study. Infants were 9-12 months old (mean age: 10 months). Infants who were premature or who had experienced frequent ear infections (three or more occurrences) were excluded from study. Each participant had normal hearing for at least the better ear at 500 – 4000 Hz as evidenced by a behavioral hearing test using visual reinforcement audiometry (VRA) in the soundfield. Middle ear function was also assessed each day of VRISD testing using the GSI TympStar (Version 2). Participants who displayed abnormal middle ear function (-150 daPa and ≤0.1 peak compliance) bilaterally were rescheduled to return for testing at a later date. If tympanometry revealed abnormal middle ear function at this later date, the participant was excluded from the study.

STIMULI

Eight stimuli were used during testing: /u/, /sa/, /a/, /i/, /ba/, /pa/, /da/, and /ga/. One speech sound contrast (u/sa) was used as a training contrast, while three contrasts were used for VRISD testing: a/i, da/ga, ba/pa. These contrasts have been used in previous VRISD studies (e.g.,Eilers et al., 1981; Rossman, 1992). The contrasts were chosen because they represent different levels of difficulty with the vowel (a/i) being easiest and the place

contrast (da/ga) being most difficult for both normal hearing children and children with sensorineural hearing loss (Boothroyd, 1984; Martinez et al., 2008).

One sound from each pair (the first from each pair listed above) served as the repeating background stimuli, while the other sound (the second from each pair) served as the change or target stimuli. Each stimulus was 500 ms in duration. Each trial consisted of three stimuli, which were separated by 1200 ms of silence, therefore making each trial 5100ms long. These stimulus parameters were consistent with previous studies that have used other infant discrimination paradigms (e.g., Kuhl, 1991; Nozza et al., 1990). While initially the order of presentation of the pairs of stimuli was randomly chosen for each subject so to eliminate an order effect, it was discovered during pilot testing that the infants tended to have more difficulty learning the VRISD task when the more difficult contrasts were presented first. This could have potentially skewed data in that some children learned the VRISD task when completing the 'easy' contrast first and went on to perform better on the more difficult contrasts whereas other infants seemingly didn't learn/remember the task until presented with the a/i contrast. Therefore, all infants first completed the a/i contrast and the presentation of the two more difficult contrasts were randomized, and followed the first contrast.

Each speech sound was recorded by an adult female in a sound-treated room using a Marshall Electronics uni-directional condenser studio-recording microphone (Model MXL 2001) with a frequency response of 30 Hz

to 20 kHz. The speaker maintained a constant microphone-to-lip distance (approximately 8 inches) and was instructed to maintain a monotone pitch by seemingly 'connecting' the speech sounds together as she produced them. The speech samples were then routed through a pre-amplifier (PreSonus Audio Electronics Tube PRE) and high pass filtered at 80 Hz. A 16 bit analog-to-digital converter (AD Instruments Power Lab/16 SP) was used to record the stimuli. Initially, stimuli were recorded at 40 kHz and then down sampled to 22050 Hz in an editing program. This software (Goldwave, v. 5.08) was also used to edit each stimulus for duration (500 ms) as well as to ramp the first and last 30 milliseconds of the sound. For consonant-vowel (CV) stimuli, each stimulus was edited so that the length of duration of the consonant was 100ms and the vowel was 400ms. Again, each stimulus' first and last 30 milliseconds was ramped to avoid any 'popping' sound that may have occurred due to sudden onset/offset of the sound. The same /a/ stimulus was used for each of the CV stimuli (i.e., it was copied and pasted to the consonant), so as to maintain consistency of the vowel sound.

## CALIBRATION

Prior to each test session, stimuli were calibrated in the soundfield using a Sinometer digital sound level meter (model JTS-1357). The sound level meter was placed at the level of the infant's head where the infant was to be seated during testing. A maximum output intensity of 60 dB SPL was

ensured using the A-weighted frequency scale and fast time weighting for each of the speech stimuli used during testing.

## INSTRUMENTATION AND TEST ENVIRONMENT

Testing was completed in a double-walled sound booth. Hearing testing was conducted using a Madsen Aurical Plus Audiometer (version 3.09) which ran on an HP Compaq dc7900 desktop computer utilizing the Microsoft Windows XP Professional operating system.

VRISD testing was performed using software created by Intelligent Hearing Systems (VideoVRA Version 2.3.0). Output from the VRISD software was run through the Madsen Aurical Plus Audiometer and played through a GSI speaker mounted to the wall of the sound booth. The video display, which was used for visual reinforcement, was located above the speaker. A button box served as the interface between the tester and the VRISD software during VRISD testing. This button box was connected to the computer via a USB port.

## VRISD TESTING

Testing was completed in two sessions: training and testing. These two sessions occurred within 3 weeks of one another. During all VRISD testing, the child was seated in a high chair or on a parent's lap in the center of the sound booth. If the infant sat in the high chair, the caretaker sat behind the child in the back of the room, out of the child's line of sight. An assistant, who

served as a distracter, sat directly in front of the child as depicted in Figure 9. Stimuli were delivered in the soundfield. The assistant used silent toys (e.g., a hand puppet) to center the child's attention when s/he was not responding to the different speech sounds.

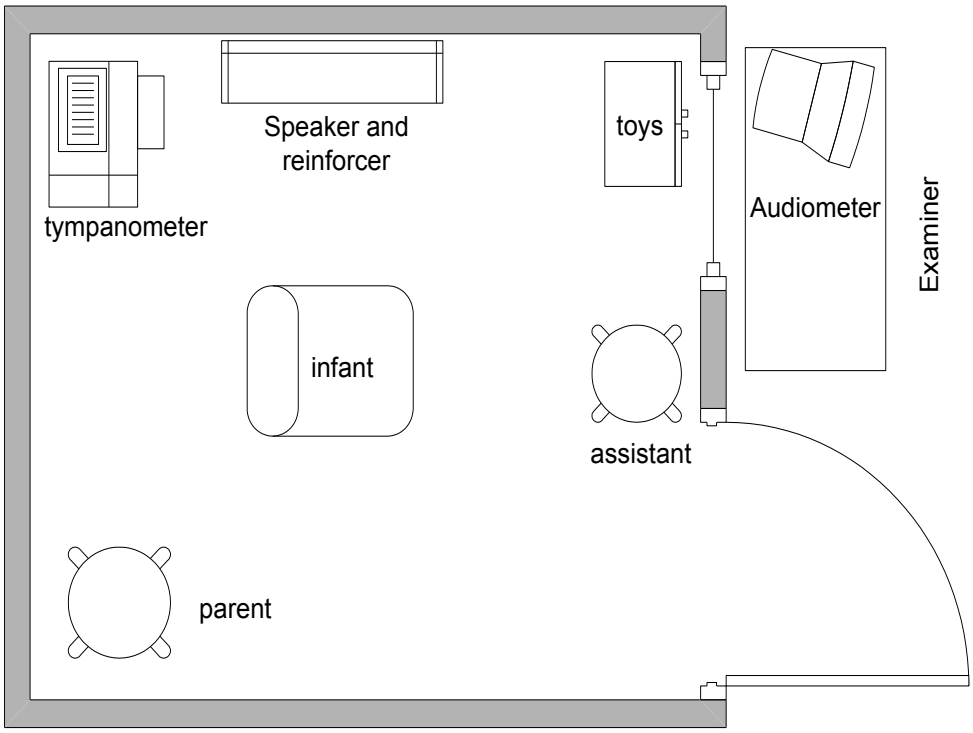


Figure 9. Sound booth configuration.

All trials were initiated by the evaluator, who was outside the sound booth in a test room. The two rooms were connected via a one-way mirror. Trials were initiated by the push of a button on a button box once the child's attention was directed toward the center. Speech sounds were presented in groups of three (i.e., /a/ /a/ /a/) at 60 dB SPL. Forty percent of the trials were control trials, with the computer program pseudo-randomly determining when each trial type was presented, the number of target trials in a row was limited to four, and the number of background trials presented consecutively to three. Thirty trials were administered during each contrast assessment. If the trial was a control trial, the background sound did not change. If it was a target trial, the novel stimulus was presented (also in a group of three, /i/ /i/ /i/), followed by a return to the background sound. The evaluator judged whether or not the child executed a head turn response during the trial. Head turns were recorded via a button press on the button box attached to the computer and the computer determined if the child's response was a correct response or a false positive. Correct responses were rewarded by visual reinforcement (i.e., the illumination of an animated video) which was located above the speaker, 90 degrees to the left of the subject

TRAINING

The first VRISD session served as a training session and took place immediately after the behavioral hearing test. The use of a training session

with a contrast different than those being assessed in actual testing has been frequently used to condition infants to the VRISD task in other studies (e.g., Eilers et al., 1984; Eilers et al., 1981). The contrast used to condition the infants to the task was u/sa. This contrast was chosen for two reasons: 1) the two stimuli were very different from one another, which would serve well for conditioning purposes, and 2) these stimuli were not stimuli in the VRISD testing being completed in this study. The training contrast was administered in a manner similar to that of the test contrasts, with one sound serving as the background stimuli while the other served as the target stimuli, but with two distinct differences. During training, the parent, assistant, and tester could hear the stimuli being played. That is, they were not blinded to the sound being presented to the infant. Also, the reinforcer was illuminated by the tester during change trials, even if the infant did not respond on his/her own. The assistant was encouraged to show the infant the reinforcer if the infant did not turn towards it after it turned on in order to help condition the child. The parent and assistant were instructed to give the infant positive feedback (i.e., clapping, saying “Good job”, etc.) when they saw a head turn followed by illumination of the reinforcer.

TESTING

This session presented all three of the stimulus pairs that were to be assessed for performance. a/i, ba/pa, and da/ga. The a/i contrast was tested first, followed by either ba/pa or da/ga, which were randomly chosen. The

infant was given a short break between testing of the different contrasts, during which the infant, parent, and assistant exited the sound booth and went into a nearby room. While this occurred, the experimenter changed the stimuli for the next test contrast. Both the parent and the assistant wore headphones playing music in the sound booth throughout the testing session to help eliminate any cueing and alleviate bias by blinding them to the sounds being heard by the infant. Again, the parent and assistant were instructed to give the infant positive feedback (i.e., clapping, saying “Good job”, etc.) when they saw a head turn followed by illumination of the reinforcer.

## ANALYSIS

The same methods of performance analysis that were used to assess pseudo subject performance were used with data obtained from the infants. These methods included: percent correct, proportion correct, discrimination index, signal detection theory, and the two criterion methods: seven-out-of-eight and binomial criterion.

## SUMMARY OF METHODS

In summary, initially, simulated data was created for 15 pseudo subjects so that performance analysis methods could be explored. This resulted in the development of a new method of performance analysis: a binomial probability- based criterion. Then 15 infants completed VRISD testing. At the first test session, each subject received a hearing test to

confirm normal hearing sensitivity for at least the better ear. Then a VRISD training contrast was administered using u/sa. At the second visit, discrimination of three speech sound contrasts (a/i, ba/pa, and da/ga) was assessed. The two visits occurred within three weeks of one another. Each test session took approximately 45 minutes. Both the parent and the assistant wore headphones and listened to music during VRISD testing to ensure that the sounds being played through the speaker were not audible to them in effort to reduce bias. Thirty trials were administered for each contrast. Results obtained for the three speech sound contrasts was then analyzed using the six different performance assessment measure: percent correct, proportion correct, discrimination index, signal detection theory, a seven correct responses out of eight consecutive trials criterion, and a binomial probability-based criterion.

# CHAPTER 4: RESULTS

## Description of Subjects

Fifteen subjects ages 9 – 12 months (6 males, 9 females; mean age: 10 months) completed testing. Consent for testing was received for two additional subjects, but they did not complete testing; the first due to scheduling conflicts, the second due to middle ear pathology. This yields a 12% attrition rate, which is quite low in comparison to many VRISD studies (e.g., Bohn & Polka, 2001; Eilers et al., 1981; Trehub et al., 1981). Demographic information was obtained for each subject's parents. All infants were from Caucasian families with parents who had college or advanced degrees. English was the primary language spoken in the homes of all subjects. All but one subject received and passed a hearing screening at birth; one infant was home birthed and did not receive a hearing screening.

## Experimenters

All participants were tested by the same experimenter, a pediatric audiologist. An assistant was in the sound booth with the parent and infant at all times. Persons who acted as assistants were either audiologists or graduate clinicians in audiology who had previous experience with pediatric testing. Test sessions were videotaped so that inter-rater reliability could be assessed.

Krippendorff's alpha was used to assess inter-rater reliability of head turn data videotaped during data acquisition. Krippendorff's alpha is a

measure of inter-rater reliability that takes inter-rater agreement due to chance into consideration (Hayes & Krippendorff, 2007; Krippendorff, 2007). The author and two additional observers watched taped test sessions and counted the total number of head turns made by infants during the test session. A Krippendorff's alpha value of 1.0 indicates perfect reliability whereas a value of 0.0 reveals chance reliability. A value of 0.8 or greater is considered sufficient inter-rater reliability (Krippendorff, 2004). Krippendorff's alpha for this study indicates very good reliability ($\alpha = 0.9846$), which implies that the three raters agreed on the occurrence of a head turn the majority of the time and that only 2% of rater responses were due to chance.

Two questions were addressed in this study: 1) How do normal hearing infants perform on 3 contrasts?, 2) What is the most appropriate statistic for determining a stopping criterion? The remainder of this chapter is organized according to these two questions.

*How do normal-hearing infants perform on the three contrasts?*

While overall performance on the three contrasts was assessed via methods previously discussed (percent correct, proportion correct, *d'*, DI, 7-out-of-8 criterion, and binomial probability criterion), only the measure of percent correct will be discussed here because this scoring method has been the most commonly used in published studies (each individual infant's performance analyzed in each of the performance analysis methods appears in the Appendices). Percent correct was defined as the overall percent correct

on all 30 trials. Performance expressed as mean percent correct for the group of infants is listed for each of the contrasts in Table 10.

| Contrast | Percent Correct (Group Mean) |
|---|---|
| a/i | 80.5% (SD = 10.6) |
| ba/pa | 66.7% (SD = 16.2) |
| da/ga | 58.2% (SD = 14.8) |

Table 10. Mean group performance expressed in percent correct and standard deviations for each of the three speech contrasts.

A one-sample *t*-test was completed to determine if the group's performance differed significantly from chance performance. Mean performance for the a/i contrast indicated performance better than chance ($t = 11.1$, $p < 0.0001$, $df = 14$), indicating that the group of infants successfully discriminated the contrast. Individual performance ranged from 53% to 93% (mean = 80.5%). Figure 10 depicts performance for each of the subjects on the a/i contrast.

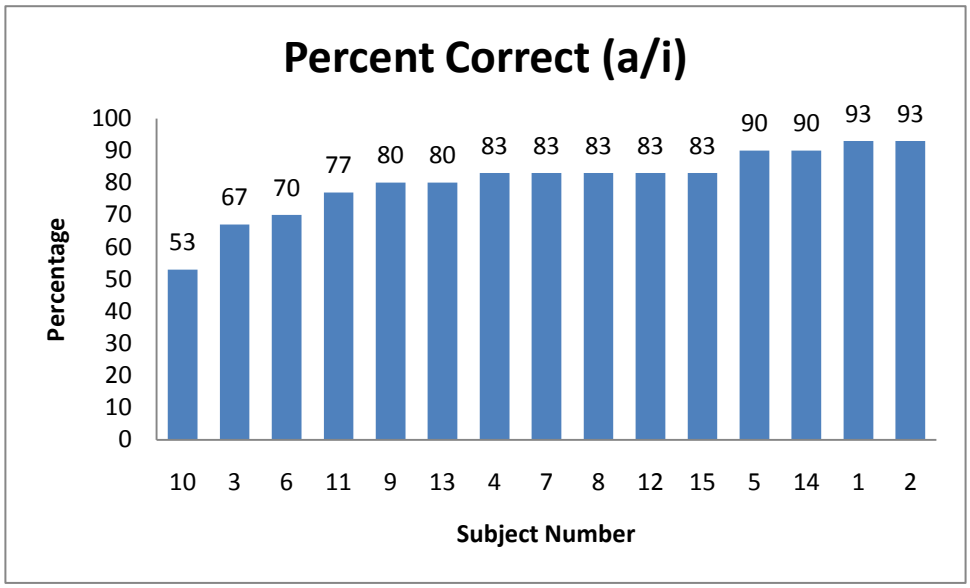


Figure 10. Individual subject performance measured in percent correct on the a/i contrast, in order from worst to best performance.

A one-sample *t*-test revealed that mean performance was better than chance for the ba/pa contrast as well ($t = 4.0$, $p = 0.001$, $df = 14$), suggesting the group successfully discriminated the contrast. Individual performance on this contrast ranged from 37% to 87% (mean = 66.7%). Figure 11 shows performance for each of the subjects on the ba/pa contrast.

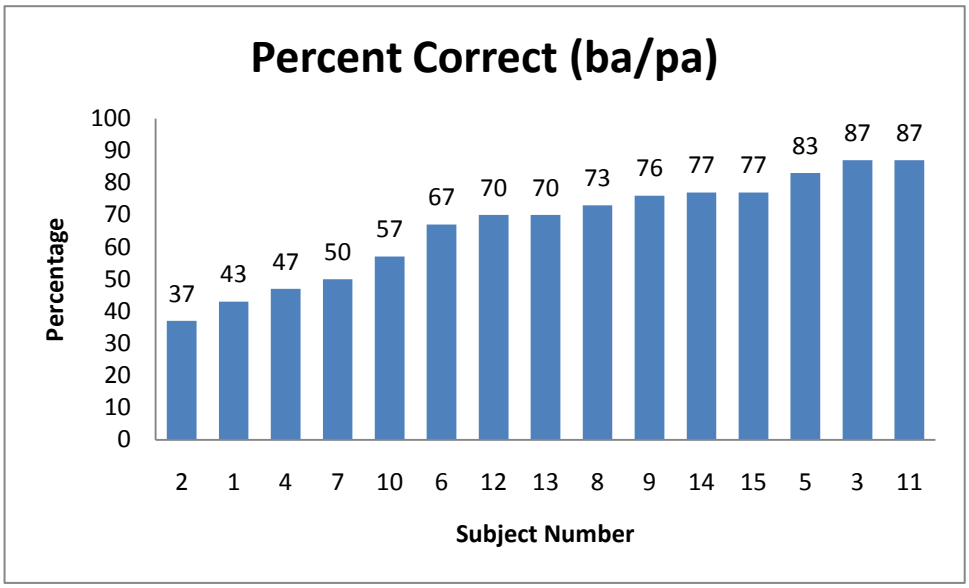


Figure 11. Individual subject performance measured in percent correct on the ba/pa contrast, in order from worst to best performance.

Mean performance for the da/ga contrast did not differ significantly from chance performance as measured by a one-sample *t*-test ($t = 2.13$, $p = 0.051$, $df = 14$), indicating that, as a group, the infants did not discriminate the contrast. Individual performance ranged from 37% to 83% (mean = 58.2%). Figure 12 depicts performance for each of the subjects on the da/ga contrast.

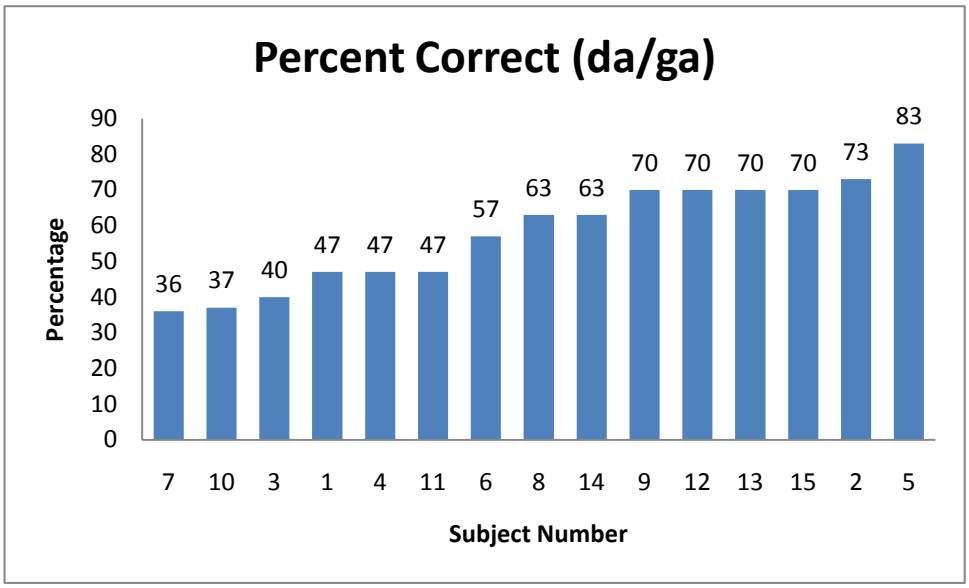


Figure 12. Individual subject performance measured in percent correct on the da/ga contrast, in order from worst to best performance.

These results suggest that the infants, as a group, successfully discriminated the a/i and ba/pa contrasts, but did not discriminate the da/ga contrast. Performance on the contrasts follows the hypothesized hierarchy of difficulty, with a/i being the easiest and da/ga being the most difficult.

The means of the group's performance on each contrast were compared to one another using one-way ANOVA, which was adjusted for correlation between contrasts. Tukey HSD results indicate that performance on the a/i contrast was significantly better than performance on either the ba/pa contrast ($p = 0.05$, $df = 2$) or the da/ga contrast ($p = 0.01$, $df = 2$). There was no significant difference between performance on the ba/pa and da/ga contrasts. Note that the trends in performance on the three contrasts can be found in all the performance assessment measures addressed in this study.

Results of performance were also analyzed using a one-way ANOVA to investigate differences in performance based on sex. No significant difference was found in the mean performance (measured in percent correct) of males and females on any of the contrasts (a/i: F = .59, p = 0.454; ba/pa: F = 1.54, p = 0.236; da/ga: F = 1.91, p = 0.189; df = 1). This is consistent with previous studies of infant speech discrimination that demonstrate no significant difference in the discrimination abilities of males versus females (e.g., Rossman, 1992).

Performance on the contrasts was also analyzed to determine if subject age was related to performance using a one-way ANOVA. No significant difference in performance on any contrast was found between the 9, 10, 11 and 12 month olds (a/i: F = 1.06, p = 0.403; ba/pa: F = 1.45, p = 0.281; da/ga: F = 1.01, p = 0.423; df = 3).

Power analysis was also performed post hoc using G*Power power analysis software (Version 3.1.2) (Faul, 2009). Power was calculated using the mean percent correct scores. Ideally, power is greater than 0.80, which suggests that if a difference in performance between the two contrasts exists, that difference would most likely be seen in the outcomes. Power was strongest for the a/i – da/ga contrast comparison (0.99), indicating that if a difference in mean performance on the two contrasts exists, this study would have probably identified it – and it did. Power was moderately strong (0.76) for the a/i – ba/pa contrast comparison (a difference in mean performance between these two contrasts was seen in this study) and only moderate

(0.51) for the ba/pa – da/ga contrast comparison (no difference in mean performance between these two contrasts was identified in this study). This relatively low level of power for the ba/pa – da/ga contrast comparison may mean that a Type II error was made – there may indeed have been a difference in the mean performance between these two contrasts, but it could not be identified in this study. Assessing more subjects on this contrast may increase the power, which would help to determine if a difference in mean performance truly exists.

*What is the best measure to use for a criterion-based stopping rule?*

Results for each subject were assessed via 6 methods: percent correct, proportion correct, discrimination index, signal detection theory (d'), 7/8 criterion, and binomial probability criterion. While individual performance varied widely, subjects generally performed best on the a/i contrast and worst on the da/ga contrast regardless of the measure of performance used.

The following tables report correlations between the different performance analysis measures for the discrimination results of each contrast tested. As can be seen in the tables, the different measures of performance analysis correlate highly with one another, indicating that they come to similar conclusions as to whether the infant successfully discriminates the contrasts.

A/i contrast. Table 11 depicts correlations between the six performance analysis measures on the a/i contrast. Pearson correlations between any and all of the analysis measures are significant at the alpha < 0.01 level, indicating

that if an infant scored well on the a/i contrast via one performance analysis measure, he/she also scored well via the other measures. Note that only one performance analysis measure does not have perfect correlation with the binomial probability measure: the 7-out-of-8 criterion. This means that the two methods of analysis did not always reach the same decision as to whether or not an infant could discriminate the contrast. On six occasions (out of 75 possible instances), the 7-out-of-8 criterion indicated results that differed from those of the binomial probability criterion. One example is Subject 6, who successfully discriminated the a/i contrast when using all analysis measures except the 7-out-of-8 criterion. In this case, the infant did not meet the criterion because the trials on which he responded incorrectly (Trials 5, 8, 12, 15, 17, 20, 25, 28, and 30) were spaced so that the 7-out-of-8 criterion could not be met, even though his performance throughout the testing of this contrast was quite strong (see this subject's cumulative distribution function for this contrast on p. 161).

**Correlations**

| | | % Correct | Proportion Correct | Discrimination Index | d' | 7-out-of-8 | Binomial Prob Crit |
|---|---|---|---|---|---|---|---|
| % Correct | Pearson Correlation | 1 | 1.000(**) | 1.000(**) | 1.000(**) | .681(**) | 1.000(**) |
| | Sig. (2-tailed) | | .000 | .000 | .000 | .005 | .000 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| Proportion Correct | Pearson Correlation | 1.000(**) | 1 | 1.000(**) | 1.000(**) | .681(**) | 1.000(**) |
| | Sig. (2-tailed) | .000 | | .000 | .000 | .005 | .000 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| Discrimination Index | Pearson Correlation | 1.000(**) | 1.000(**) | 1 | 1.000(**) | .681(**) | 1.000(**) |
| | Sig. (2-tailed) | .000 | .000 | | .000 | .005 | .000 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| d' | Pearson Correlation | 1.000(**) | 1.000(**) | 1.000(**) | 1 | .681(**) | 1.000(**) |
| | Sig. (2-tailed) | .000 | .000 | .000 | | .005 | .000 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| 7-out-of-8 | Pearson Correlation | .681(**) | .681(**) | .681(**) | .681(**) | 1 | .681(**) |
| | Sig. (2-tailed) | .005 | .005 | .005 | .005 | | .005 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| Binomial Prob Crit | Pearson Correlation | 1.000(**) | 1.000(**) | 1.000(**) | 1.000(**) | .681(**) | 1 |
| | Sig. (2-tailed) | .000 | .000 | .000 | .000 | .005 | |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |

Table Caption

** Correlation is significant at the 0.01 level (2-tailed).

Table 11. Correlations between performance analysis measures for performance results on the a/i contrast.

Ba/pa contrast. Table 12 shows correlations between the six performance analysis measures on the ba/pa contrast. Pearson correlations between any and all of the analysis measures are significant at least at the 0.05 level, indicating that if an infant scored well on the ba/pa contrast via one performance analysis measure, he/she also scored well via the other measures. The binomial probability criterion measure correlates perfectly with percent correct and *d'* measures and while it highly correlates to the other measurements of performance, conclusions drawn from the binomial probability criterion measure are not always the same as those drawn from the 7-out-of-8 criterion, DI and proportion correct measures. Again, an example is Subject 6. Subject 6 demonstrated discrimination of the ba/pa contrast via percent correct, proportion correct, *d'*, and the binomial probability criterion (see Subject 6's performance profile on p. 157). A closer look at Subject 6's trial-by-trial performance shows that he responded incorrectly to only four of the first 15 trials, but then missed four of the next six trials. His performance improved for the last nine trials, responding incorrectly to two of the nine. This poor performance in the middle of the test brought down scores overall - only the binomial probability criterion was not affected (the subject had reached the criterion on Trial 8). Like his performance on the a/i contrast (discussed on p. 79), the trials on which the subject responded incorrectly were spaced so that the 7-out-of-8 criterion could not be met.

**Correlations**

| | | % Correct | Proportion Correct | Discrimination Index | d' | 7-out-of-8 | Binomial Prob Crit |
|---|---|---|---|---|---|---|---|
| % Correct | Pearson Correlation | 1 | .756(**) | .577(*) | 1.000(**) | .577(*) | 1.000(**) |
| | Sig. (2-tailed) | | .001 | .024 | .000 | .024 | .000 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| Proportion Correct | Pearson Correlation | .756(**) | 1 | .764(**) | .756(**) | .327 | .756(**) |
| | Sig. (2-tailed) | .001 | | .001 | .001 | .234 | .001 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| Discrimination Index | Pearson Correlation | .577(*) | .764(**) | 1 | .577(*) | .667(**) | .577(*) |
| | Sig. (2-tailed) | .024 | .001 | | .024 | .007 | .024 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| d' | Pearson Correlation | 1.000(**) | .756(**) | .577(*) | 1 | .577(*) | 1.000(**) |
| | Sig. (2-tailed) | .000 | .001 | .024 | | .024 | .000 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| 7-out-of-8 | Pearson Correlation | .577(*) | .327 | .667(**) | .577(*) | 1 | .577(*) |
| | Sig. (2-tailed) | .024 | .234 | .007 | .024 | | .024 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| Binomial Prob Crit | Pearson Correlation | 1.000(**) | .756(**) | .577(*) | 1.000(**) | .577(*) | 1 |
| | Sig. (2-tailed) | .000 | .001 | .024 | .000 | .024 | |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table 12. Correlations between performance analysis measures for performance results on the ba/pa contrast.

Da/ga contrast. Table 13 depicts correlations between the six performance analysis measures on the da/ga contrast. Pearson correlations between any and all of the analysis measures are significant at least at the 0.05 level, indicating that if an infant scored well on the da/ga contrast via one performance analysis measure, he/she also scored well via the other measures. That said, the binomial probability criterion did not correlate with perfect agreement with any of the other measures of performance. This indicates that there were many instances where the binomial probability criterion resulted in a different conclusion as to the discrimination ability of an infant than the other measures of performance. The binomial probability criterion did reach the same conclusions as the other measures of performance analysis for 10 of the 15 subjects (all the measures agreed that the infant either did or did not discriminate the contrast). Again, Subject 6 is an example of the disagreement between performance measures. The binomial probability criterion and proportion correct measurements agreed that this subject successfully discriminated the contrast. While the proportion correct measure did indicate discrimination of the contrast, the score was not indicative of strong performance: the subject achieved a score of 0.61 (a score of 0.56 or greater indicated successful discrimination). The binomial probability

criterion was met after Trial 12 – he missed three of the first 12 trials. The subject's performance then decreased; he responded incorrectly to 10 of the remaining 18 trials presented. In cases such as this, the use of a criterion to measure performance makes the most sense: a criterion takes advantage of the novelty of the task and, therefore, the child's attention to the task.

**Correlations**

| | | % Correct | Proportion Correct | Discrimination Index | d' | 7-out-of-8 | Binomial Prob Crit |
|---|---|---|---|---|---|---|---|
| % Correct | Pearson Correlation | 1 | .873(**) | .764(**) | 1.000(**) | .875(**) | .756(**) |
| | Sig. (2-tailed) | | .000 | .001 | .000 | .000 | .001 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| Proportion Correct | Pearson Correlation | .873(**) | 1 | .667(**) | .873(**) | .764(**) | .866(**) |
| | Sig. (2-tailed) | .000 | | .007 | .000 | .001 | .000 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| Discrimination Index | Pearson Correlation | .764(**) | .667(**) | 1 | .764(**) | .873(**) | .577(*) |
| | Sig. (2-tailed) | .001 | .007 | | .001 | .000 | .024 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| d' | Pearson Correlation | 1.000(**) | .873(**) | .764(**) | 1 | .875(**) | .756(**) |
| | Sig. (2-tailed) | .000 | .000 | .001 | | .000 | .001 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| 7-out-of-8 | Pearson Correlation | .875(**) | .764(**) | .873(**) | .875(**) | 1 | .661(**) |
| | Sig. (2-tailed) | .000 | .001 | .000 | .000 | | .007 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| Binomial Prob Crit | Pearson Correlation | .756(**) | .866(**) | .577(*) | .756(**) | .661(**) | 1 |
| | Sig. (2-tailed) | .001 | .000 | .024 | .001 | .007 | |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Table 13. Correlations between performance analysis measures for performance results on the ba/pa contrast.

Performance on each contrast as measured by the binomial probability criterion was assessed for differences in mean performance on the contrasts as well as for differences in mean performance based on age and/or sex, as was done for performance as measured by percent correct. This was done to ensure that the end results were similar to one another regardless of the performance analysis measure used.

Results of performance using the binomial probability criterion measure were assessed using one-way ANOVA which was adjusted for correlation between contrasts. Tukey HSD results indicated that, on average, infants performed significantly better on the a/i contrast than da/ga contrast ($p = 0.01$; $df = 2$). While results for the a/i versus ba/pa comparison approached a significant alpha level, results failed to meet it, indicating that there was no significant difference in performance between the a/i and ba/pa contrasts nor was a significant difference present for performance between the ba/pa and da/ga contrasts. These results are consistent with the pattern observed when analysis was done using percent correct.

ANOVA was done to investigate possible differences in performance based on sex and age. Consistent with results obtained in percent correct, no significant difference was found for the performance of males versus the performance of females (a/i: $F = 1.0$, $p = 0.33$; ba/pa: $F = 0.01$, $p = 0.89$; da/ga: $F = 0.19$, $p = 0.66$; $df = 1$) nor was a difference found for performance based on age (a/i: $F = 0.86$, $p = 0.48$; ba/pa: $F = 1.03$, $p = 0.41$; da/ga: $F = 2.01$, $p = 0.17$; $df = 3$).

Group performance for the three speech sound contrasts is depicted in Table 14. A "*" next to a measure indicates successful discrimination of the contrast to which it corresponds according to that method's manner of determining successful discrimination. Percent correct, proportion correct, DI, and d' were calculated using performance on all 30 trials, as is typically done in published literature on infant speech discrimination, and the scores and outcomes are revealed in Table 14. Performance when using a criterion measure would typically be described as "pass" or "not pass", but for purposes of comparison, the average number of trials necessary until infants reached criterion is depicted. When using a criterion procedure to assess performance, it appears that each contrast was successfully discriminated in fewer than 30 trials, although it took more trials to demonstrate successful discrimination of the ba/pa and da/ga contrasts than was necessary for the a/i contrast.

| | Percent correct | Proportion correct | Discrimination index | d' | 7/8 criterion | Binomial criterion |
|---|---|---|---|---|---|---|
| a/i | 80.5%*[a] (SD=10.6) | 0.80*[a] (SD=0.09) | 0.67* (SD=0.2) | 1.88* (SD=0.6) | 12.4* (SD=8) | 10.1* (SD=5.7) |
| ba/pa | 66.7%*[c] (SD=16.2) | 0.69*[a] (SD=0.1) | 0.45 (SD=0.3) | 1.1* (SD=0.8) | 19.7* (SD=9.4) | 15.9* (SD=10.4) |
| da/ga | 58.2% (SD=14.9) | 0.6*[b] (SD=0.1) | 0.3 (SD=0.3) | 0.72 (SD=0.8) | 23.3* (SD=8) | 19.9* (SD=8.4) |

* = successful discrimination according to that method's manner of determining success
[a] = $p < 0.0001$
[b] = $p < 0.001$
[c] = $p < 0.005$

Table 14. Mean group performance on each of the three speech contrasts as measured by percent correct, proportion correct, discrimination index, signal detection theory (d'), 7-out-of-8 criterion, and binomial probability criterion.

Mean performance on the a/i contrast indicated successful discrimination regardless of the performance analysis measure used. When assessing criterion options for performance assessment of the a/i contrast, it is clear that the average subject could demonstrate successful discrimination of the a/i contrast in a fraction of the number of trials that were actually presented. Via the 7-out-of-8 criterion, successful discrimination was demonstrated in an average of 12 trials. Recall though, that achieving 7-out-of-8 trials in a row correct in more than 10 trials is not statistically significant; such performance could potentially occur by chance. Use of the binomial probability criterion yields successful discrimination of the a/i contrast in only 10 trials with statistical significance ($p<0.05$), thereby permitting testing to be complete in approximately 1/3 the time it would normally take to present 30 trials.

Figure 13 depicts mean performance on the a/i contrast by comparing performance measured in percent correct to performance measured by the binomial probability criterion. Because a low number of trials in a criterion measure indicates better performance, the binomial probability criterion was transformed (multiplied by -1) so that a comparison of performance between percent correct and binomial probability criterion could be made. This method has been used by Liu et. al. in a previous study (Liu, Kuhl, & Tsao, 2003). It can be seen that, in general, those infants who scored lower using the percent correct method also needed more trials to reach the binomial probability criterion.

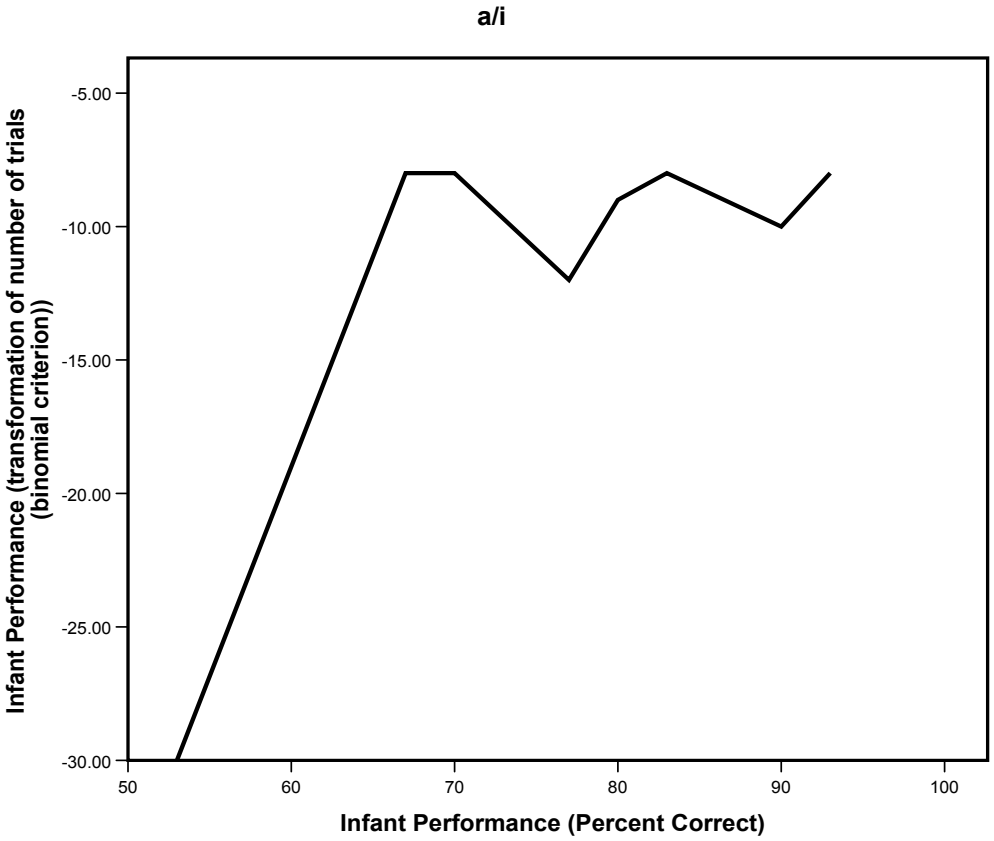


Figure 13. Performance in percent correct versus binomial probability criterion (transformed) for the a/i contrast.

Mean performance on the ba/pa contrast yields mixed results: the two criterion methods of performance assessment indicate successful discrimination of the contrast as do percent correct, proportion correct, and signal detection theory, while the discrimination index method does not. When looking at individual performance scores, 11 infants successfully

discriminated the contrast as evidenced by percent correct and 12 infants demonstrated discrimination via proportion correct scores. Only 10 infants demonstrated discrimination when using the binomial probability criterion to assess performance. Two subjects (Subjects 7 and 10) had relatively low percent correct and proportion correct scores. Neither of these subjects demonstrated discrimination of the ba/pa contrast via the binomial probability criterion. Subject 7's trial-by-trial data show that she missed 15 of the 30 trials; she responded incorrectly to seven of the first 15 trials, and to eight of the last 15 trials. Closer inspection of Subject 10's trial-by-trial data show that this infant responded incorrectly to 10 of the first 15 trials. Her performance then improved – she missed only three of the final 15 trials.

Like the a/i contrast, both criterion methods show that infants, in general, could demonstrate successful discrimination of the ba/pa contrast in fewer than 30 trials. The 7-out-of-8 criterion could be met in an average of 19 trials, but again, this performance could occur by chance. The binomial probability criterion demonstrates that discrimination of the ba/pa contrast could be met with statistical significance ($p<0.05$) in an average of 15 trials, decreasing test time by approximately half.

Figure 14 depicts performance on the ba/pa contrast by comparing the percent correct and binomial probability criterion methods of performance analysis. Again, the number of trials necessary to reach criterion was transformed by multiplying by -1 to indicate that a lower number of trials was consistent with better performance. Those infants who had lower percent

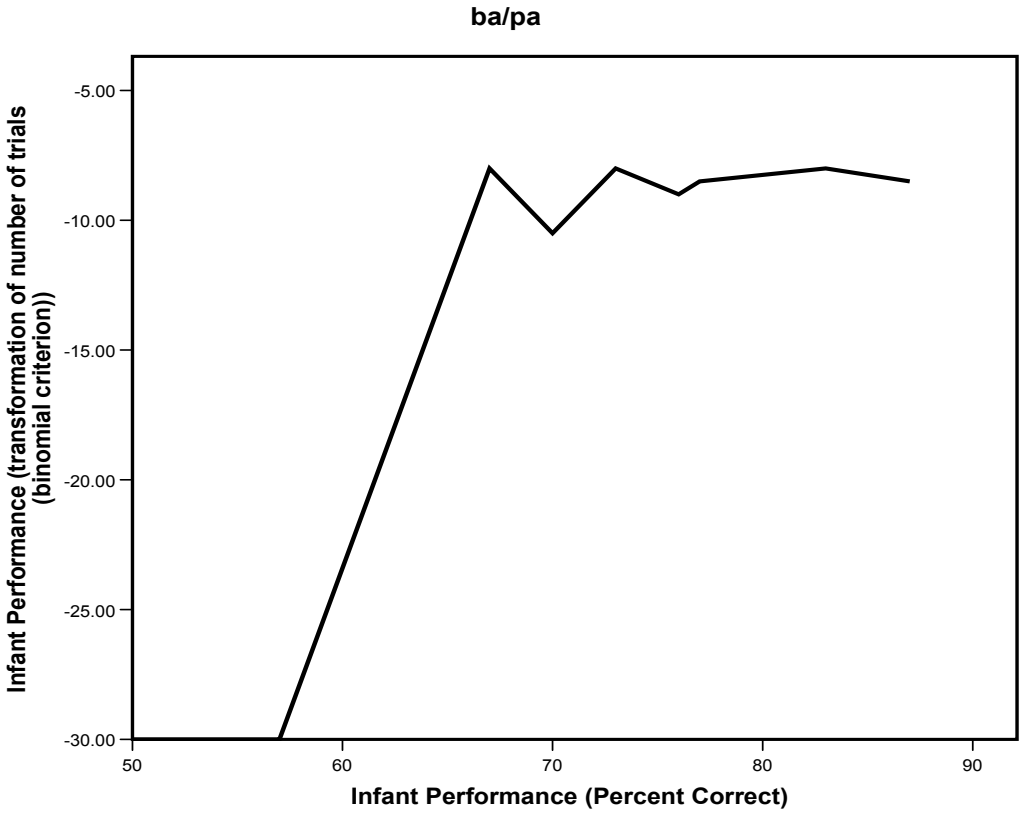


Figure 14. Performance in percent correct versus binomial probability criterion (transformed) for the ba/pa contrast.

correct scores again required more trials for the binomial probability criterion method.

Performance of subjects on the da/ga contrast was worse than performance on either the a/i or the ba/pa contrast, regardless of the methods of performance analysis used. Eight subjects successfully discriminated the contrast when the percent correct measure of performance analysis was used. Ten infants successfully discriminated the contrast when performance

was measured using the binomial probability criterion, eight of these 10 being the eight infants who successfully discriminated this contrast according the percent correct measure. Two infants who successfully discriminated this contrast only did so when the binomial probability criterion was used.

Again, both criterion methods show successful discrimination of the da/ga contrast in less than 30 trials: the 7-out-of-8 criterion in 23 trials, the binomial probability criterion in 19 trials. Achieving 7-out-of-8 correct trials in a row could again occur by chance, whereas the binomial probability criterion performance is statistically significant ($p < 0.05$). Figure 15 portrays performance on the da/ga contrast by comparing the percent correct and binomial probability criterion measures of performance analysis. The number of trial necessary to reach criterion was again transformed by multiplying by -1 so that a lower number of trials was indicative of better performance. Once again, in general, those infants with lower percent correct scores required more trials to meet the binomial probability criterion, or did not meet the criterion at all.

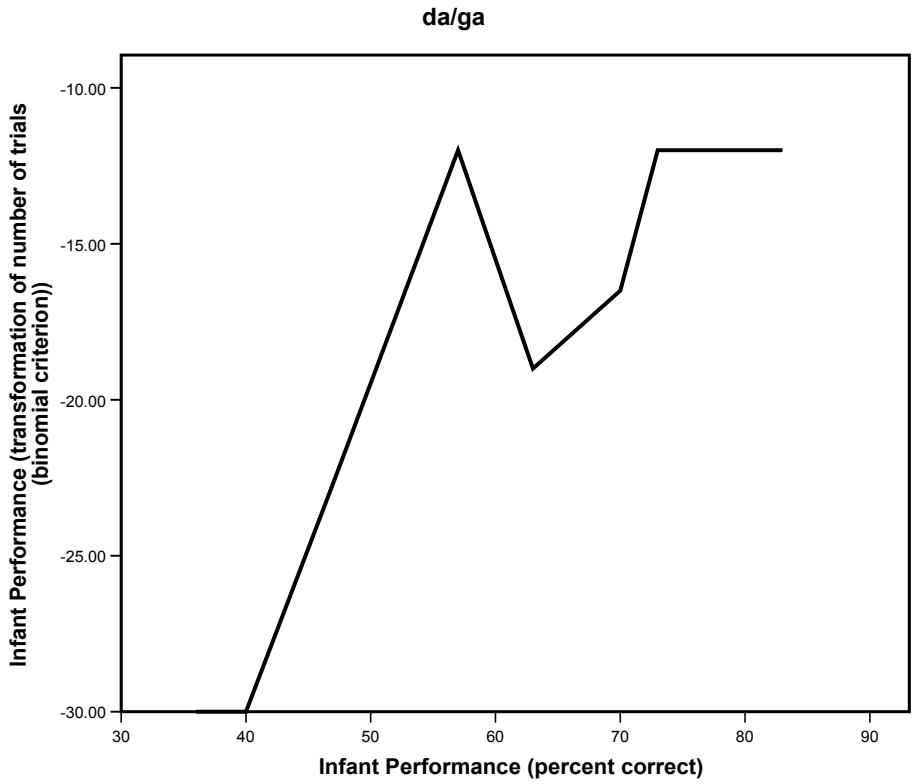


Figure 15. Performance in percent correct versus binomial probability criterion (transformed) for the ba/pa contrast.

Table 15 shows the number of subjects who successfully reached the binomial probability criterion for each of the speech sound contrasts. To help illustrate the number of trials necessary to meet the criterion, “number of trials” was divided into groups: 8-10 (recall that eight was the minimum number of trials in which the criterion could be reached), 11-15, 16-20, 21-25, 26-29, and 30 (that is, these infants did not meet criterion/did not demonstrate discrimination).

| Number of Trials | a/i | ba/pa | da/ga |
|---|---|---|---|
| 8 – 10 | 12 | 9 | 2 |
| 11-15 | 2 | 1 | 3 |
| 16-20 | 0 | 0 | 4 |
| 21-25 | 0 | 0 | 1 |
| 26-29 | 0 | 0 | 0 |
| 30 (did not reach criterion for discrimination) | 1 | 5 | 5 |

Table 15. Number of subjects who reached binomial probability criterion for each of the speech sound contrasts.

Twelve infants demonstrated successful discrimination of the a/i contrast within 8 to 10 trials, whereas nine out of the 15 infants demonstrated discrimination of the ba/pa contrast within that same number of trials, and only two of the 15 infants could complete the da/ga contrast in 8 to 10 trials. Five infants did not reach the criterion for either the ba/pa or the da/ga contrast. This will be discussed further in the Discussion section of this paper (p. 106).

Figure 16 depicts individual subject performance using the binomial probability criterion method of performance analysis on the a/i contrast. Scores are revealed as the last trial number necessary to meet the criterion. It is easy to see that the majority of infants demonstrated successful discrimination of the a/i contrast in significantly fewer than the 30 trials that were administered.

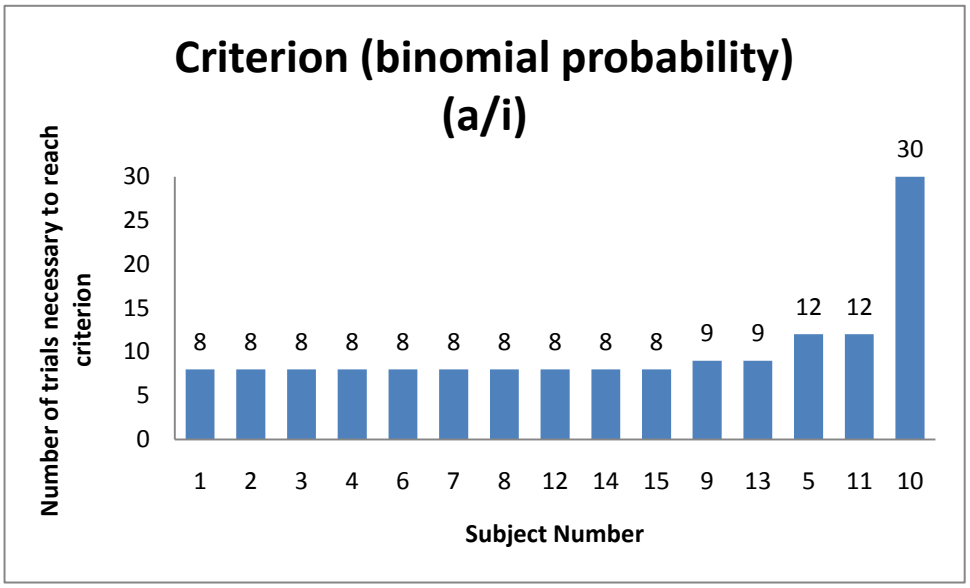


Figure 16. Subject performance on the a/i contrast using the binomial probability criterion. Scores are revealed as the last trial number necessary to meet criterion, in order from the fewest number of trials necessary to the greatest.

Only one infant did not demonstrate discrimination of the a/i contrast in any analysis measure. This subject frequently responded with a head turn to the change of the target stimuli when it returned to the background stimuli (seven such "off responses" were counted during the a/i contrast). Had these responses occurred within the response interval, this subject would have shown successful discrimination of the contrast. Because these responses occurred outside of the response window, they did not count as correct responses. More about "off responses" will be addressed in the discussion section of this paper (p. 112).

Figure 17 depicts performance of each infant on the ba/pa contrast, presenting the number of trials necessary to reach the binomial probability

criterion. Ten of the infants demonstrated discrimination of the contrast in 12 trials or less. Subject 10 (the infant who did not successfully discriminate a/i) again had many "off-responses" that, had they occurred within the response interval, would have shown successful discrimination.

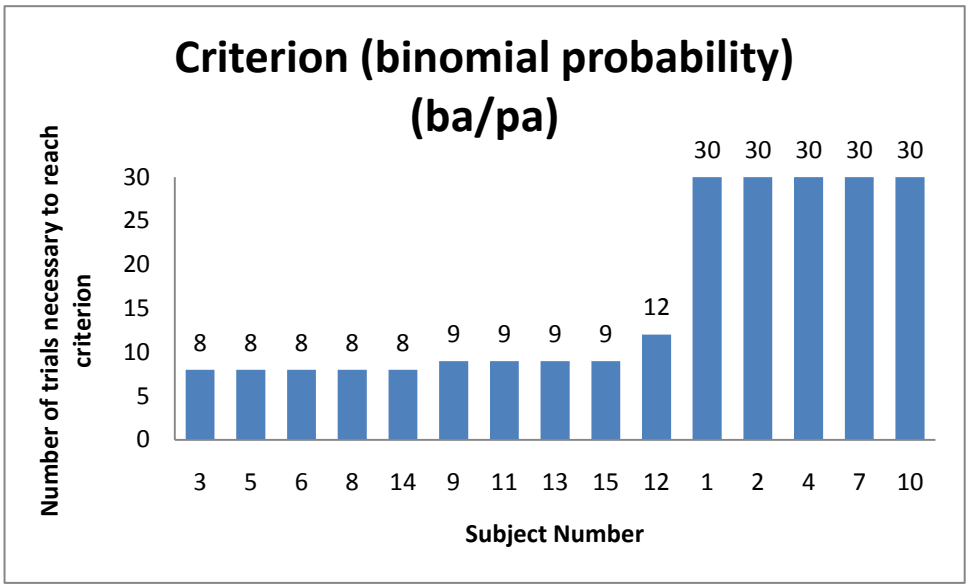


Figure 17. Subject performance on the ba/pa contrast using the binomial probability criterion. Scores are revealed as the last trial number necessary to meet criterion, in order from the fewest number of trials necessary to the greatest.
.

Figure 18 shows performance of each infant on the da/ga contrast, again depicting the number of trials necessary to meet the criterion. Performance on this contrast was highly variable in comparison to performance on the other contrasts. Ten of the 15 infants met the binomial probability criterion in 22 trials or fewer. Note that it took an average of 14.8 trials for infants to reach criterion on the da/ga contrast whereas it took an

average of 8.8 trials to reach criterion on the ba/pa contrast and 8.7 trials to reach criterion on the a/i contrast.

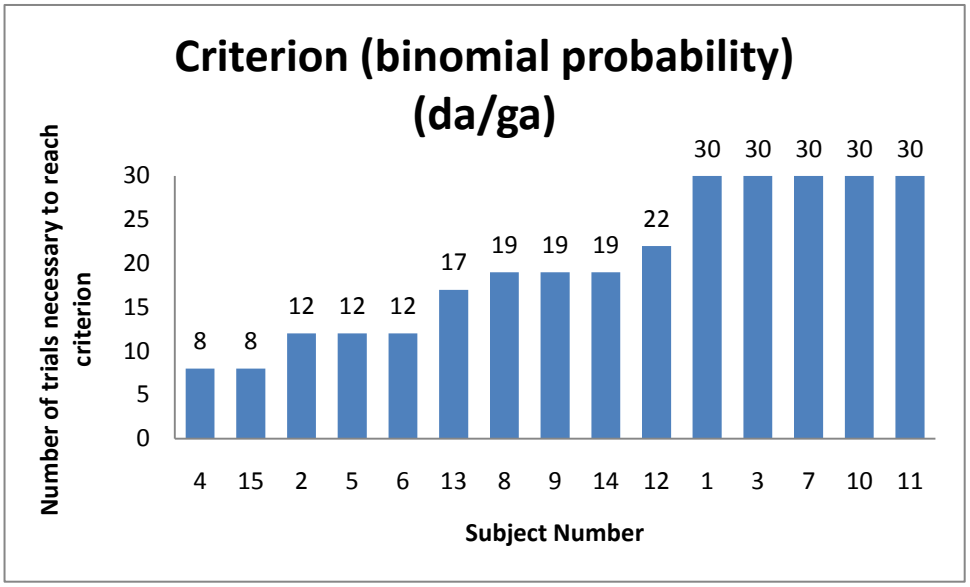


Figure 18. Subject performance on the da/ga contrast using the binomial probability criterion. Scores are revealed as the last trial number necessary to meet criterion, in order from the fewest number of trials necessary to the greatest.

.

Individual Results and Comparisons. A more in depth look into the results of individual infants yields further evidence that a criterion-based stopping point, and the binomial probability criterion in particular, is the best way to analyze individual infant performance/successful discrimination of a contrast. See Appendices for detailed information regarding each infants' performance.

All methods of performance analysis agreed that Subjects 1, 2, 3, 4, 5, 7, 8, 9, 11, 12, 13, 14, and 15 successfully discriminated the a/i contrast (see Table 16). The use of the two criterion measures would have enabled testing

to be finished in less than half of the amount of time it took for the 30 trials to be completed. One three occasions, while the two criterion methods did agree on results, the binomial probability criterion was reached before the 7-out-of-8 criterion. In one case (Subject 11), the binomial probability criterion was reached in 12 trials whereas it took 22 trials for the subject to reach the 7-out-of-8 criterion. Recall that reaching this criterion in 22 trials could occur by chance. Subject 10 did not discriminate the a/i contrast as evidenced by all methods of analysis except proportion correct. Her proportion correct score was 0.58, which was the minimum score that had a *t* value with statistical significance. Only the 7-out-of-8 criterion method did not reach the conclusion that Subject 6 successfully discriminated the a/i contrast. Subject 6 did not meet the criterion because the trials on which he responded incorrectly (Trials 5, 8, 12, 15, 17, 20, 25, 28, and 30) were spread out so that the 7-out-of-8 criterion could not be met.

Individual Performance - a/i

| Subject | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| 1 | 93 | 0.93 | 0.89 | 2.98 | 8 | 8 |
| 2 | 93 | 0.89 | 0.83 | 2.56 | 8 | 8 |
| 3 | 67 | 0.71 | 0.44 | 1.38 | 8 | 8 |
| 4 | 83 | 0.82 | 0.72 | 1.89 | 8 | 8 |
| 5 | 90 | 0.89 | 0.83 | 2.56 | 10 | 12 |
| 6 | 70 | 0.71 | 0.5 | 1.10 | 30 | 8 |
| 7 | 83 | 0.79 | 0.72 | 2.14 | 8 | 8 |
| 8 | 83 | 0.82 | 0.72 | 1.90 | 8 | 8 |
| 9 | 80 | 0.79 | 0.67 | 1.64 | 11 | 9 |
| 10 | 53 | 0.58 | 0.22 | 0.54 | 30 | 30 |
| 11 | 77 | 0.75 | 0.61 | 1.40 | 22 | 12 |
| 12 | 83 | 0.82 | 0.72 | 1.90 | 8 | 8 |
| 13 | 80 | 0.81 | 0.67 | 1.73 | 11 | 9 |
| 14 | 90 | 0.89 | 0.83 | 2.56 | 8 | 8 |
| 15 | 83 | 0.82 | 0.72 | 1.90 | 8 | 8 |

Table 16. Individual performance for the a/i contrast. Highlighted scores are results that indicate successful discrimination of the contrast.

All performance analysis methods agreed that Subjects 5, 9, 11, 14, and 15 discriminated the ba/pa contrast with success, and all analysis methods also agreed that Subjects 1, 2 and 4 did not discriminate the contrast (see Table 17). Percent correct, proportion correct, *d'*, and the binomial probability criterion reached the conclusion that Subjects 6 and 8 could discriminate the ba/pa contrast, but the 7-out-of-8 criterion and DI did not. Both subjects' responses to the trials on which they responded were spread out so that the 7-out-of-8 criterion could not be met. Both subjects had a few false positives which, in effect, reduced the number of hits for each

subject when the DI was calculated, which resulted in lower discrimination index scores than had the false positives not occurred.

Individual Performance - ba/pa

| Subject | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| 1 | 43 | 0.53 | 0.06 | 0.15 | 30 | 30 |
| 2 | 37 | 0.46 | 0.06 | -0.54 | 30 | 30 |
| 3 | 87 | 0.84 | 0.78 | 0.71 | 8 | 8 |
| 4 | 47 | 0.52 | 0.11 | 0.09 | 30 | 30 |
| 5 | 83 | 0.81 | 0.72 | 2.50 | 8 | 8 |
| 6 | 67 | 0.68 | 0.44 | 0.96 | 30 | 8 |
| 7 | 50 | 0.57 | 0.17 | 0.62 | 30 | 30 |
| 8 | 73 | 0.74 | 0.56 | 1.26 | 30 | 8 |
| 9 | 76 | 0.78 | 0.61 | 1.56 | 11 | 9 |
| 10 | 57 | 0.63 | 0.28 | 0.95 | 21 | 30 |
| 11 | 87 | 0.83 | 0.78 | 2.36 | 14 | 9 |
| 12 | 70 | 0.7 | 0.5 | 1.02 | 17 | 12 |
| 13 | 70 | 0.72 | 0.5 | 1.25 | 15 | 9 |
| 14 | 77 | 0.78 | 0.61 | 1.56 | 8 | 8 |
| 15 | 77 | 0.78 | 0.61 | 1.56 | 13 | 9 |

Table17. Individual performance for the ba/pa contrast. Highlighted scores are results that indicate successful discrimination of the contrast.

All performance analysis methods agreed that Subjects 2, 5, 9, 12, 13, and 15 successfully discriminated the da/ga contrast and that Subjects 3, 7, 10 and 11 did not (Table 18).

Individual Performance - da/ga

| Subject | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| 1 | 47 | 0.56 | 0.11 | 0.53 | 30 | 30 |
| 2 | 73 | 0.74 | 0.56 | 1.26 | 13 | 12 |
| 3 | 40 | 0.3 | 0 | -0.33 | 30 | 30 |
| 4 | 47 | 0.55 | 0.11 | 0.42 | 30 | 8 |
| 5 | 83 | 0.85 | 0.72 | 2.15 | 13 | 12 |
| 6 | 57 | 0.61 | 0.28 | 0.69 | 30 | 12 |
| 7 | 36 | 0.42 | -0.06 | -0.53 | 30 | 30 |
| 8 | 63 | 0.64 | 0.39 | 0.71 | 19 | 19 |
| 9 | 70 | 0.72 | 0.5 | 1.25 | 20 | 19 |
| 10 | 37 | 0.46 | -0.06 | -0.54 | 30 | 30 |
| 11 | 47 | 0.53 | 0.11 | 0.20 | 30 | 30 |
| 12 | 70 | 0.72 | 0.5 | 1.25 | 22 | 22 |
| 13 | 70 | 0.75 | 0.5 | 1.75 | 14 | 17 |
| 14 | 63 | 0.66 | 0.39 | 0.81 | 30 | 19 |
| 15 | 70 | 0.72 | 0.5 | 1.25 | 9 | 8 |

Table 18. Individual performance for the da/ga contrast. Highlighted scores are results that indicate successful discrimination of the contrast.

Of particular interest for this contrast are the results for Subjects 4 and 6 – both reached the binomial probability criterion in relatively few trials, but did not score well on the other methods of performance analysis. Subject 4 reached the binomial probability criterion for the da/ga contrast in only 8 trials. Figure 19 reveals Subject 4's performance in percent correct and binomial probability over all 30 trials. Through the eighth trial, the subject scored 75% correct (proportion correct = 0.75), which indicates that performance to this point of testing is better than chance. Based on this figure, it appears that this subject's performance deteriorated over time; the subject may have become bored or uninterested in the task. If indeed this was the case, the use of a

criterion method to analyze her performance would be vital – the use of 30 trials to determine her ability to discriminate yields falsely negative information.

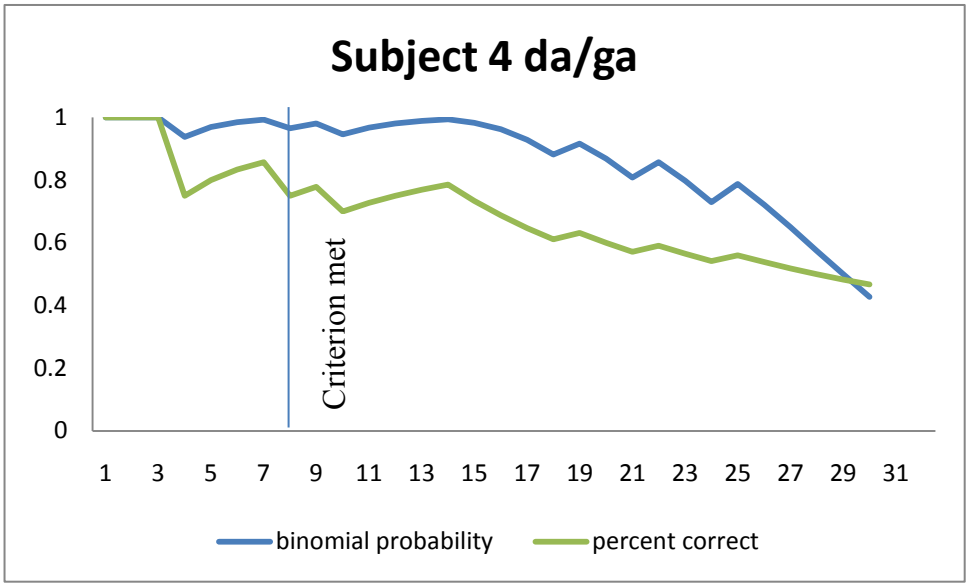


Figure 19. Subject 4's performance on the da/ga contrast as measured by percent correct and binomial probability over 30 trials.

Subject 6's performance over the 30 trials as measured by percent correct and binomial probability is represented in Figure 20. This subject's performance does not decrease over the session as markedly as Subject 4, but does decrease slightly. At Trial 12 (at which point the binomial probability criterion has been met), the subject has achieved 75% correct (proportion correct = 0.76), which is better than chance performance. The slight decrease in performance over the test session may be indicative of an infant who has grown tired of the task, making a criterion measure the ideal measure to assess performance.

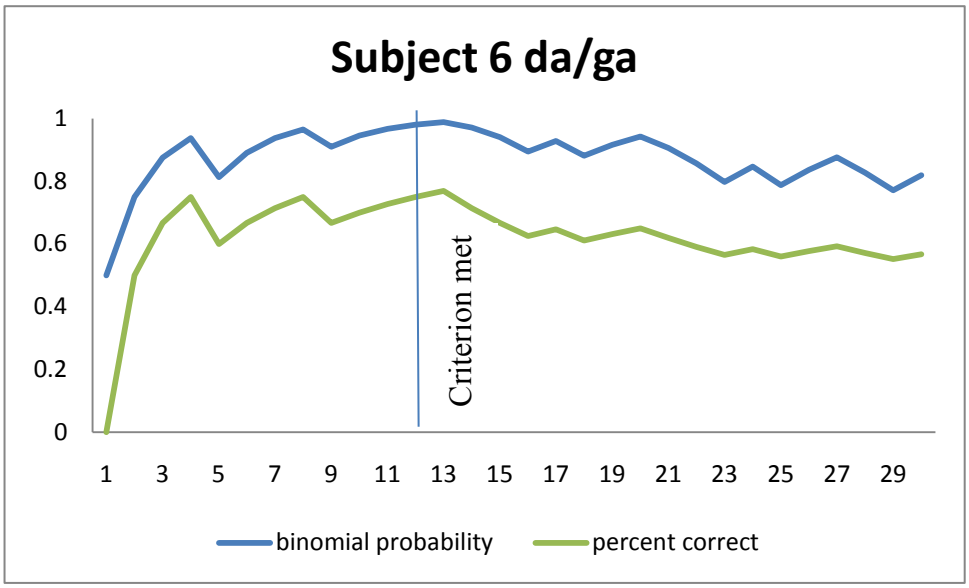


Figure 20. Subject 6's performance on the da/ga contrast as measured in percent correct and binomial probability.

Table 19 outlines the performance of infants who reached the criterion for each of the contrasts assessed. Information provided for each contrast includes the minimum and maximum number of trials that were necessary for infants to reach the binomial probability criterion. In parentheses is the number of infants who achieved the criterion in that particular number of trials. The last row presents the total number of infants who reached criterion, indicating successful discrimination of the contrast, regardless of the number of trials necessary to reach the criterion.

| | a/i | ba/pa | da/ga |
|---|---|---|---|
| Minimum number of trials necessary to demonstrate discrimination (number of infants who achieved criterion in the minimum number of trials) | 8 (10) | 8 (5) | 8 (2) |
| Maximum number of trials necessary to demonstrate discrimination (number of infants who achieved criterion in the maximum number of trials) | 12 (2) | 12 (1) | 22 (1) |
| Total number of infants who reached criterion | 14 | 10 | 10 |

Table 19. Minimum and maximum number of trials necessary to demonstrate successful discrimination. The number in parentheses is the number of infants who achieved that criterion.

According to this information, some infants were able to show successful discrimination for each of the contrasts in only 8 trials. For the a/i and ba/pa contrasts, no infant needed more than 12 trials to demonstrate successful discrimination, whereas up to 22 trials were necessary for an infant to successfully discriminate the da/ga contrast.

A criterion method to assess performance on the VRISD task allows testing to be complete once the criterion is met. It takes advantage of the novelty of the contrast and VRISD task on those first few trials, making the most of an infant's attention span and capabilities. This is critical for those infants who become habituated or fatigued with the VRISD task and cannot continue to perform at a high level over 30 trials. A criterion based on probability assesses the likelihood that events (in the case of VRISD, these events are correct responses) occur randomly by looking at the pattern of responses. Based on results of the current study, it appears that the use of the binomial probability criterion is successful in determining discrimination of the speech contrasts tested while achieving confidence in the results via statistical significance and providing the capability to reduce test time.

DISCUSSION – CHAPTER 5

This study addressed two primary questions using the VRISD procedure: How do normal-hearing infants perform on three contrasts? What is the most appropriate statistic for determining a stopping criterion?

Abilities of 15 infants with normal hearing to discriminate the phoneme contrasts a/i, ba/pa, and da/ga were evaluated. Results of discrimination ability for each of the contrasts were compared with one another. A hierarchy of difficulty for the contrasts was hypothesized based on previous literature, with the a/i contrast being the easiest and the da/ga contrast being the most difficult (Boothroyd, 1984; Martinez et al., 2008). As a group, results obtained via percent correct indicated that the infants successfully discriminated the a/i (mean score = 80.5%) and ba/pa (mean score = 66.7%) contrasts, but not the da/ga contrast (mean score = 58.2%). This pattern of discrimination abilities for the contrasts was seen across all performance analysis measures. These results are consistent with the hypothesized hierarchy of difficulty for discrimination.

While all measures of performance demonstrated this trend of discrimination ability, the use of a binomial probability criterion-based performance measure yielded more individual subjects with the ability to discriminate all three contrasts. Use of 30 trials to assess discrimination ability doesn't appear to enhance performance, but can, in actuality, show deterioration in performance. This decrease in performance may be due to boredom or fatigue and indicates that a fixed number of trials is not

necessarily the optimal indicator of an infant's ability to discriminate. Results analyzed using the commonly utilized percent correct method and those analyzed using the binomial probability criterion method came to similar conclusions (i.e., the a/i contrast was easiest, there was no difference in performance based on age or sex), but the binomial probability criterion method allowed testing to stop once the infant demonstrated successful discrimination of the contrast rather than continuing testing to a pre-determined number of trials. The use of this criterion-based method to assess performance on VRISD contrasts ensures an accurate conclusion regarding an infant's ability to discriminate is drawn, especially in cases where the infant demonstrates ability to discriminate during the beginning of the test session but becomes bored with the task as it continues to 30 trials. It also reduces the test time and could lead to the potential of the testing more contrasts than could be completed with the use of more traditional measures of performance analysis, most of which require the use of up to 30 trials.

Five infants did not reach the criterion for either the ba/pa or the da/ga contrast. The fact that this many infants with normal hearing in our study had difficulty demonstrating discrimination of these contrasts shows that VRISD testing and performance must be further investigated before the VRISD technique can be used as a measure of discrimination ability, especially for children with hearing loss. A few aspects to consider include boredom, response type, and practice/learning. Given that these infants had normal hearing, it is likely that they could indeed hear the change from the

background to the target stimuli. It cannot be known exactly why these babies did not demonstrate discrimination on this task. Perhaps the infant was bored or tired. One infant in this study was an “off responder” – she turned her head when the target sound changed back to the background sound (more on this is discussed on page 112). Perhaps the infant truly could not discriminate the contrast. Of note is that infants do not achieve 100% performance, even on easy contrasts, on speech perception tasks so perhaps the fact that he/she did not demonstrate discrimination for a more difficult contrast is simply as aspect of this (see page 110 for more discussion on this topic). Some VRISD studies incorporate training on a contrast prior to testing discrimination of it – perhaps this training and practice would be enough to facilitate successful discrimination during testing (more on training and practice is discussed on page 112). One study brought infants back for repeat testing, which demonstrated that some infants seemed to learn the task over time (Uhler, 2008b). Perhaps the infants who did not demonstrate ability to discriminate in the current study would have been able to do so had they returned for additional testing.

Since parameters of the stimuli and test procedures were fairly similar, results of the current study’s results on the ba/pa contrast can be compared to those of Eilers et al. (1981). Eilers and colleagues included the ba/pa contrast in their study on discrimination of voice onset time in infants using the VRISD procedure (1981). The infants in their study were younger than infants in the present study (6-8 months versus 9-12 months, respectively).

The average discrimination index score for the ba/pa contrast which consisted of naturally spoken stimuli in the 1981 study was 0.375, indicating successful discrimination of the contrast. Infants in our study also successfully discriminated the contrast, with a mean DI of 0.45. There was no significant difference between performance of the group in the current study and Eilers et al.'s subjects, as evidenced by a *t* -test ($t = 0.74$, $p = 0.465$; $df = 21$). The fact that the two groups of infants performed similarly in spite of the two studies varying somewhat in their procedures, Eilers et al.(1981) used a different training technique, stimuli were recorded by a male speaker, and stimuli were 473 ms long) is evidence that discrimination of the ba/pa contrast is fairly robust and that the VRISD technique, even with small variations in parameters, can be used successfully to assess infant speech discrimination in 6-12 month olds.

The use of a criterion-based measure to assess discrimination ability in infants may yield more information than simply an infant's ability to discriminate. Liu et al. (2003) and Tsao, Liu, and Kuhl (2004) suggest that the number of trials necessary for an infant to reach criterion can be regarded as an index of differences in infants' speech perception skills. No data is presented in these studies to support this suggestion. Recall that infants in the current study reached the binomial probability criterion in an average of 8.7 and 8.8 trials for the a/i and ba/pa contrasts, respectively, and an average of 14.8 trials for the da/ga contrast. The fact that some infants reach criterion on a particular contrast earlier than others, or that typical infants can achieve

the criterion in fewer trials for one contrast than for another contrast, may be indicative of differences in perceptual skills, but there is also evidence that differences in performance are indicative of cognitive development (Lalonde & Werker, 1995). Further study is necessary to understand the implications of the number of trials necessary for an infant to reach a criterion.

Based on the range of the number of trials necessary for individual infants in this study to reach the criterion (see Table 19), it can be seen that requiring infants who do not demonstrate discrimination of a contrast within a particular window of time to complete all 30 trials may be fruitless; it simply prolongs testing of a contrast for which the infant cannot/does not discriminate. Further investigation is necessary to develop a stopping criterion to cease the test procedure when it is evident that an infant will not/cannot show evidence of discrimination for that contrast.

While VRISD is a useful paradigm to assess speech discrimination abilities in infants and has potential for clinical use, there are some important aspects that must be considered when using the VRISD paradigm: even infants with normal hearing do not score 100% correct, negative results must be interpreted with caution, types of responses accepted during testing need to be considered/clarified, practice and learning effects must be considered, how to use VRISD with infants who have hearing loss, and the determining what contrasts should be included in testing. These issues need to be addressed regardless of which of the infant speech perception assessment measures is used, as none of these issues have been addressed in infant

speech perception testing. Each of these items is discussed in more detail below:

Failure of infants to reach 100% correct - Throughout many of the VRISD studies cited in this paper, the failure of infants to reach 100% performance is acknowledged. No subject in this study performed with 100% accuracy. Should this inability affect how performance on VRISD interpreted? Even on an “easy” contrast, most infants, even those who seem to be in an ideal state for testing (i.e., not fussy, not overly interested in the distracter while also not overly interested in the reinforcers themselves, etc.), will not achieve a score of 100% correct. Nozza explains VRISD as a task of vigilance in which the subject maintains attention to the background in order to detect a signal (Nozza, 1987). He points out that, according to the vigilance model, the decrease in correct responses occurs due to changes in response bias rather than changes in ability to detect a signal.

Although an infant may perform perfectly for a period of time, in general, performance does not reach 100% for infants (Nozza et al., 1990). The fact that infants typically do not achieve 100% correct may be interpreted as a problem with the percent correct and proportion correct measures, although maximum performance at less than 100% (or 1.0 if using the proportion correct measure) is not uncommon in infants (see Eilers et al., 1984; Nozza et al., 1991; Nozza et al., 1990 and other VRISD results), even in VRA testing. It is probable that infants are less likely than adults to achieve

100% correct performance due to false positives and misses that are due to inattention or other testing factors (Nozza et al., 1990).

Interpretation of negative results - Caution needs to be used when drawing conclusions about a child who does not demonstrate successful discrimination on a contrast. Initially it may appear that the child could not detect the difference between the two stimuli tested, but this may not be a valid or correct conclusion. The child may have grown bored of the task, been distracted by hunger or a dirty diaper, or simply tired. One way we have found to help differentiate between a child who truly cannot detect the differences of a contrast from a child who is simply bored is to return to an easier contrast that the child has successfully completed in the past. If the child continues to discriminate that contrast with success, it gives the tester some confidence that the child indeed could not detect the differences in the contrast for which they did not pass. This method to confirm that inability to demonstrate the ability to discriminate two speech sounds has also been used by Werker and Tees (1984). In our pilot studies, when an infant with hearing loss did not demonstrate successful discrimination of a contrast, it was recommended the parent and early interventionist 'play' with the sounds in that contrast with the child and then return for testing two to three weeks later. Many times, the child would return and be able to discriminate the contrast with success. Of course, audibility of the sounds in the contrast should be ensured as well.

Types of Accepted Responses - While a head turn is the criterion for response in the conditioned head turn task, it has been observed on occasion

that, rather than turning his/her head towards the reinforcer (or sometimes in addition to it), the infant repeated the stimuli. While for the purposes of this study, those times that an infant repeated stimuli but did not turn towards the reinforcer would not count as correct, those involved with this study agreed that in the clinical setting, this type of response would be accepted.

There were also instances when the infant turned to the reinforcer when the target stimuli changed back to background stimuli. In the clinic, this is sometimes seen during VRA testing. Typically, the audiologist will count this "off response" as correct. Given the limits of the software (the reinforcer would only come on if the child responded within 5 seconds of the onset of the trial), these off responses could not be reinforced and were not counted as correct responses. Again, those involved in this study agreed that in the clinical setting, more often than not, this type of response would be accepted.

Practice and Learning – Practice and learning have an impact on infants' performance on the VRISD task. Some infants seem to learn the task quickly and easily while others require more exposure and even assistance (i.e., drawing the infant's attention to the reinforcer) to learn the task. During data collection in pilot studies, it was noted that infants were quite good at generalizing the VRISD task to new contrasts. Therefore, it was decided not to use conditioning trials for each contrast assessed in this study, but to instead administer a training contrast. This method of conditioning infants to the VRISD task has been used in other studies (e.g., Eilers et al., 1984; Eilers et al., 1981). The training contrast was administered at the first visit (as

described on p. 67), after the hearing test. The three contrasts were then tested at the second visit. During initial data collection, the presentation order of the three contrasts was randomized. It was noted at that time, that an infant who was tested on the a/i contrast first appeared to perform better on the two more difficult contrasts whereas an infant who was tested on either of the other two contrasts first appeared to have a more difficult time with the task. Given the methodology planned for this study, once this difference in performance was observed, it was decided to administer the easier a/i contrast first. This easier contrast seemed to serve as a sort of training contrast in and of itself.

Some previous VRISD studies have used training contrasts (e.g., Eilers et al., 1981)while others have used conditioning trials (e.g., Nozza, 1987). Conditioning trials consist of the two sounds being assessed in the contrast, but include an intensity cue to help condition the infant to the task. For example, if /u/ serves as the background sound and /a/ serves as the target, /u/ is presented at the intensity testing will take place at (in our testing, 60dB), while /a/ is presented at a higher intensity (sometimes 10dB greater). This intensity cue can be used in addition to the formant differences in the stimuli to help the infant learn the task. Typically, once the infant correctly responds to two presentations of the target with the intensity cue, the intensity cue is reduced. This is repeated until the background and the target stimuli are of the same intensity. Once this occurs, testing begins.

No study has been conducted to investigate if one method of training is better than another. Perhaps what works well for one infant would not work well for another. That said, infants in the current study may have performed as well (or perhaps even better) on the more difficult contrasts even if they were assessed first had they had been given conditioning trials using those speech sounds with an intensity cue.

In visual reinforcement audiometry (VRA), the pediatric audiologist tends to begin testing using monitored live speech to obtain a speech awareness threshold. Conditioning the infant to the task of VRA using speech stimuli tends to be more effective and faster (for most children) than using tonal stimuli. If the pediatric audiologist begins conditioning for VRA testing using tonal stimuli but the infant doesn't condition fairly quickly, the tester typically will try conditioning using speech stimuli instead. In summary, in the audiology clinic, the pediatric audiologist tends to condition an infant to testing by using 'easy' stimuli. Utilizing the a/i contrast first may serve that same purpose.

Using VRISD with Infants with Hearing Loss - The data for this dissertation was obtained from infants with normal hearing sensitivity. While our team has conducted pilot studies that confirm that VRISD can be successfully performed by infants and toddlers with hearing loss, we have not yet conducted a study to assess the performance of those infants and toddlers using the binomial probability criterion. That said, two infants with hearing loss have recently been tested on the a/i contrast using the same

parameters as those used in the study presented in this paper, except that the maximum number of trials presented was 20.

One infant (age 10 months) assessed using VRISD was diagnosed with auditory dys-synchrony via auditory brainstem response (ABR) and otoacoustic emissions (OAE) testing following newborn hearing screening. Her behavioral audiograms at ages 7-10 months indicated responses in the mild to moderate hearing loss range. Her parents noted that she seemed to have “good hearing days” and “bad hearing days”. They reported that she was having a good hearing day on the day of VRISD testing. Testing was performed without amplification in the soundfield at 60dB SPL. This baby successfully discriminated the a/i contrast with 80% accuracy. Assessment of performance using the binomial probability criterion indicates that she would have demonstrated successful discrimination in 8 trials, similar to many of the normal hearing babies that were tested for this dissertation. See Figure 21 for the cumulative distribution function for this contrast, which shows that this infant demonstrated discrimination of the contrast from the beginning of the test, and continued successfully discriminating the sounds throughout the test session.

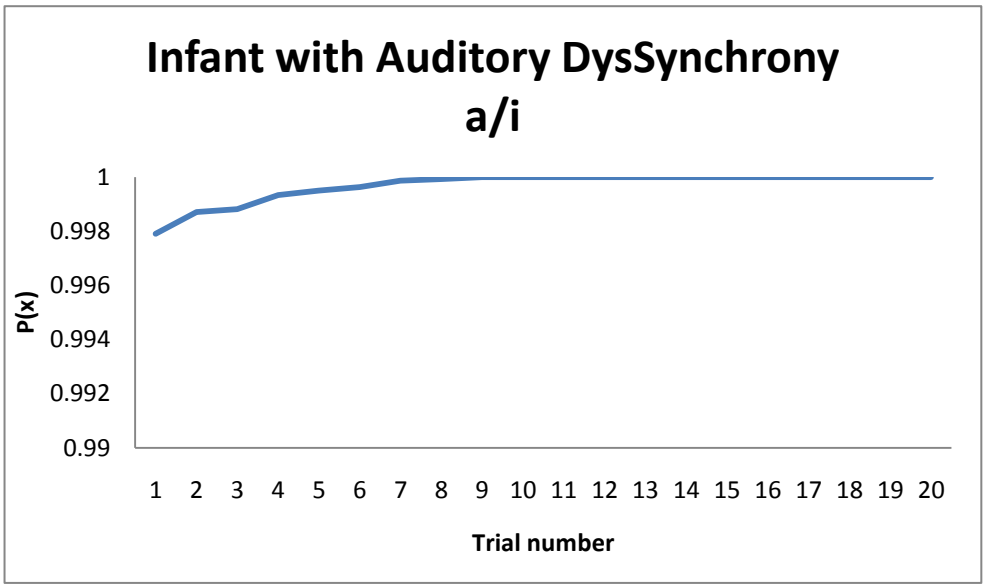


Figure 21. Cumulative distribution function for infant with hearing loss (auditory dys-synchrony) on the a/i contrast.

The other infant was an 11 month old male diagnosed with severe-profound sensorineural hearing loss in the right ear and moderate-severe sensorineural hearing loss in the left ear via ABR after a referral from his newborn hearing screening. He has worn hearing aids since two months of age and is receiving a cochlear implant in May 2010. VRISD testing was performed in the soundfield at 60 dB SPL while the infant wore his hearing aids. His performance (expressed in both percent correct and binomial probability) is depicted in Figure 22. After 30 trials, he scored 60% correct for the a/i contrast, which does indicate successful discrimination – although barely (the minimum score necessary to infer ability to discriminate is 58%). This infant did not demonstrate discrimination of the a/i contrast via the binomial probability criterion measure. As can be seen from Figure 22, his performance was somewhat inconsistent – he appeared to be able to

discriminate the contrast fairly well during the middle of the test session, but was not able to do so consistently enough to evoke confidence in his ability to do so.

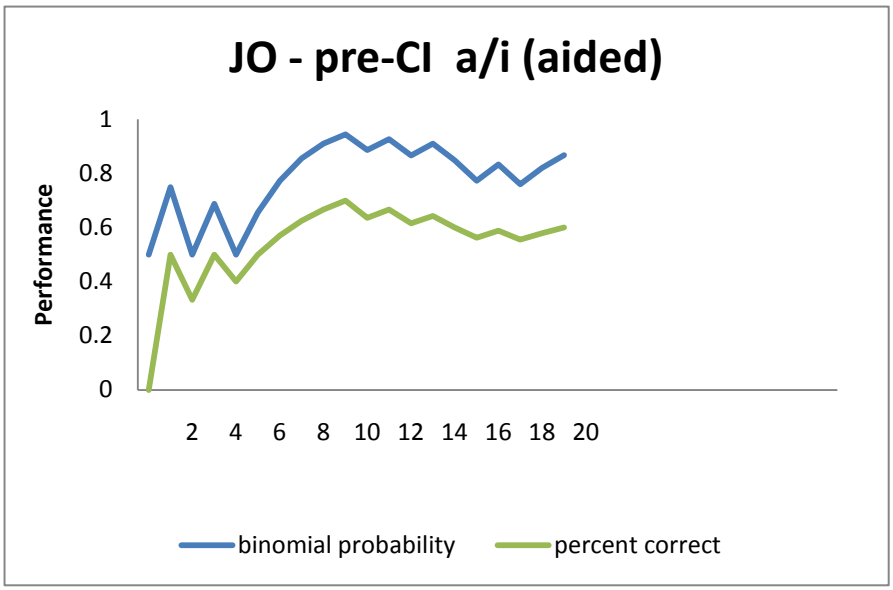


Figure 22. VRISD results for an infant with hearing loss on the a/i contrast while using his hearing aids.

Data obtained from three infants with hearing loss during an earlier pilot study were re-analyzed using the binomial probability criterion method. The VRISD parameters used in the pilot study were similar to those used in this dissertation, but there were no limits of consecutive trials of the same type. Regardless, data from all three infants with hearing loss demonstrated discrimination of the a/i contrast in 8 to 14 trials while using their amplification.

What Contrasts Should be Assessed? - As has been pointed out by Werker et al., one positive aspect of VRISD is its ability to use practically any stimuli (Werker et al., 1997). This study used three speech contrasts: a vowel

contrast (a/i), and two consonant-vowel contrasts (ba/pa and da/ga). Many different speech contrasts have been used in previous VRISD research, ranging from simple consonant-vowel contrasts (e.g., Eilers et al., 1977; Nozza, 1987) to English phonemes that have been acoustically altered (e.g., Kuhl, 1991; Trainor & DesJardins, 2002) to multisyllabic nonsense words (e.g., Eilers et al., 1984). Research has not yet been conducted to highlight which speech sound contrasts are best to include in VRISD testing or how typically developing infants perform on these contrasts. Theoretically, the "best" contrasts to use will depend on what information is sought. An investigation into the use of VRISD to determine benefit from amplification will indeed use stimuli different from an investigation into a hierarchy of difficulty of discrimination or a study of discrimination of non-native phonemes. To use VRISD in a clinical setting, a protocol (or protocols) should be developed to address clinical questions.

**Future Directions**

While VRISD holds potential for clinical use, more study is necessary to further develop, refine, and validate the VRISD procedure. Many possible areas of future research on VRISD have been discussed in this section. The binomial probability criterion appears to have potential for use as a method to measure an individual infant's performance on the task with confidence while reducing test time. The use of this criterion needs to be further investigated with the study of more infants, both hard of hearing and normal hearing, to determine if it indeed is an appropriate measure of performance for this task.

A criterion to stop testing when it is apparent that an infant will not demonstrate discrimination of a contrast should also be further investigated. The notion that the number of trials required to reach a criterion may hold additional information should also be further investigated. If the number of trials necessary to reach a criterion can indeed be used as an index of differences in infants' speech perception abilities (as suggested by Liu et al., 2003; Tsao et al., 2004) or as an indicator of cognitive development (as suggested by Lalonde & Werker, 1995), VRISD could be used not only as a test of speech perception, but also development. Study into what constitutes average performance for normally-developing infants should also be continued. The use of conditioning and training trials needs to be investigated to find the best way to condition infants to the VRISD task.

**Conclusions**

The results of this dissertation indicate that VRISD can be used with success to investigate infant speech perception. Based on the results of 15 infants who were tested on three speech contrasts (a/i, ba/pa, and da/ga), the binomial probability criterion succeeds in measuring performance on the VRISD task. Using the binomial probability criterion to determine successful discrimination of a speech contrast can reduce the number of trials by up to two-thirds, while maintaining statistical significance. The use of fewer trials would enable accurate conclusions to be drawn regarding the abilities of infants who, for whatever reason, may not stay attentive for VRISD testing that uses 30 trials. This reduction in the number of trials presented not only

reduces test time, but also permits more contrasts to be assessed in each test session.

References

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A Statistical Basis for speech sound discrimination. *Language and Speech, 46*(2-3), 155-182.

Aslin, R. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development, 52*, 1135-1145.

Bohn, O., & Polka, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *Journal of the Acoustical Society of America, 110*(1), 504-515.

Boothroyd, A. (1984). Auditory Perception of Speech Contrasts by Subjects with Sensorineural Hearing Loss. *Journal of Speech and Hearing Research, 27*, 134-144.

Boothroyd, A. (2004). *Measuring Auditory Speech-Perception Capacity in Young Children*. Paper presented at the Sound Foundation Through Early Amplification.

Bornstein, M. H., & Benasich, A. A. (1986). Infant habituation: assessments of individual differences and short-term reliability at five months. *Child Development, 57*(1), 87-99.

Bull, D., Eilers, R., & Oller, D. (1984). Infants' discrimination of intensity variation in multisyllabic stimuli. *Journal of the Acoustical Society of America, 76*(1), 13-17.

Bull, D., Eilers, R. E., & Oller, D. K. (1985). Infants' discrimination of final syllable fundamental frequency in multisyllabic stimuli. *Journal of the Acoustical Society of America, 77*(1), 289-295.

Ching, T. Y., & Hill, M. (2007). The Parents' Evaluation of Aural/Oral Performance of Children (PEACH) Scale: Normative Data. *Journal of the American Academy of Audiology, 18*(3), 220-235.

Culpepper, B., & Thompson, G. (1994). Effects of reinforcer duration on the response behavior of preterm 2year-olds in visual reinforcement audiometry. *Ear and Hearing, 15*, 161-167.

Early Hearing Detection and Intervention (EHDI) Program. (2010, February 16). Retrieved March 23, 2010, from www.cdc.gov/nbcddd/ehdi/data/htm

EHDI Legislation. (2009, February 3, 2009). Retrieved March 23, 2010, from http://www.infanthearing.org/legislative/mandates.html

Eilers, R. E., Bull, D. H., Oller, D. K., & Lewis, D. C. (1984). The discrimination of vowel duration by infants. *Journal of the Acoustical Society of America, 75*(4), 1213-1218.

Eilers, R. E., & Gavin, W. J. (1981). The evaluation of infant speech perception skills: Statistical techniques and theory development. In R. E. Stark (Ed.), *Language Behavior in Infancy and Early Childhood* (pp. 185-213). New York: Elsevier North Holland, Inc.

Eilers, R. E., Morse, P. A., Gavin, W. J., & Oller, D. K. (1981). Discrimination of voice onset time in infancy. *Journal of the Acoustical Society of America, 70*(4), 955-965.

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research, 20*(4), 766-780.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science, 171*(3968), 303-306.

Eisenberg, L. S., Martinez, A. S., & Boothroyd, A. (2007). Assessing auditory capabilities in young children. *International Journal of Pediatric Otorhinolaryngology, 71*, 1339-1350.

Faul, F. (2009). G*Power (Version 3.1.2). Kiel, Germany.

Gravel, J. S. (1989). Behavioral assessment of auditory function. *Seminars in Hearing, 10*, 216-228.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychoacoustics*. New York: John Wiley & Sons.

Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures, 1*(1), 77-89.

Houston, D. M., Pisoni, D. B., Iler Kirk, K., Ying, E. A., & Miyamoto, R. T. (2003). Speech perception skills of deaf infants following cochlear implantation: a first report. *International Journal of Pediatric Otorhinolaryngology, 67*(5), 479-495.

Jusczyk, P. W., & Luce, P. A. (2002). Speech Perception and Spoken Word Recognition: Past and Present. *Ear and Hearing, 23*(1), 2-40.

Kaplan, P. S., & Werner, J. S. (1986). Habituation, reponse to novelty, and dishabituation in human infants: tests of a dual-process theory of visual attention. *Journal of Experimental Child Psychology, 42*(2), 199-217.

Karzon, R. K. (1982). Discrimination of multisyllabic sequences by young infants. Central Institute for the Deaf, Washington University.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Krippendorff, K. (2007). Computing Krippendorff's Alpha-Reliability.

Kuhl, P. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not:. *Perception and Psychophysics, 50*(2), 93-107.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America, 100*(15), 9096-9101.

Lalonde, C. E., & Werker, J. F. (1995). Cognitive Influences on Cross-Language Speech Perception in Infancy. *Infant Behavior and Development, 18*, 459-475.

Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), F1-F10.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (Second ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Marean, G. C., Werner, L., & Kuhl, P. (1992). Vowel Categorization by Very Young Infants. *Developmental Psychology, 28*(3), 396-405.

Martinez, A. S., Eisenberg, L. S., Boothroyd, A., & Visser-Dumont, L. (2008). Assessing Speech Pattern Contrast Perception in Infants: Early Results on VRASPAC. *Otology and Neurotology, 29*, 183-188.

Miller, J. L., & Eimas, P. D. (1996). Internal structure of voicing categories in early infancy. *Perception and Psychophysics, 58*(8), 1157-1167.

Nozza, R. J. (1987). Infant Speech-sound discrimination testing: Effects of stimulus intensity and procedural model on measures of performance. *Journal of the Acoustical Society of America, 81*(6), 1928 - 1938.

Nozza, R. J. (1998). *Thresholds are not Enough: Understanding How Infants Process Speech has a Role in How We Manage Hearing Loss*. Paper presented at the Sound Foundation Through Early Amplification.

Nozza, R. J., Miller, S. L., Rossman, R. N., & Bond, L. C. (1991). Reliability and validity of infant speech-sound discrimination-in-noise thresholds. *Journal of Speech and Hearing Research, 34*(3), 643-650.

Nozza, R. J., Rossman, R. N., Bond, L. C., & Miller, S. L. (1990). Infant speech-sound discrimination in noise. *Journal of the Acoustical Society of America, 87*(1), 339-350.

Oller, D. K., Eilers, R. E., Burns, R., & Urbano, R. (1993). Perception of the stop/glide contrast in infancy. *Phonetica, 50*(1), 1-14.

Oller, D. K., Eilers, R. E., Miskiel, E., Burns, R., & Urbano, R. (1991). The stop/glide boundary shift: modelling perceptual data. *Phonetica, 48*(1), 32-56.

Olsho, L. (1987). An observer-based psychoacoustic procedure for use with young infants. *Developmental Psychology, 23*, 627-640.

Olsho, L. W., Koch, E. G., Halpin, C. F., & Carter, E. A. (1987). An observer-based psychoacoustic procedure for use with young infants. *Developmental Psychology, 23*, 627-640.

Pastore, R. E., Cawley, E. J., Berens, M. S., & Skelly, M. A. (2003). "Nonparametric" *A'* and other modern misconceptions about signal detection theory. *Psychonomic Bulletin & Review, 10*(3), 556-569.

Pegg, J., & Werker, J. (1997). Adult and infant perception of two English phones. *Journal of the Acoustical Society of America, 102*(6), 3742-3753.

Rossman, R. N. (1992). The effect of simulated mild-to-moderate high-frequency hearing loss on the discrimination of speech sounds by infants. Unpublished Dissertation. University of Pittsburgh.

Schmida, M. J., Peterson, H. J., & Tharpe, A. M. (2003). Visual Reinforcement Audiometry Using Digital Video Disc and Conventional Reinforcers. *American Journal of Audiology, 12*, 35-40.

Spence, M. J., & Moore, D. S. (2003). Categorization of Infant-directed Speech: Development from 4 to 6 Months. *Developmental Psychology, 42*(1), 97-109.

Summary of Infants Diagnosed Before 3 Months of Age ( Year 2007). (2009, August 2009). Retrieved March 23, 2010, from www.cdc.gov/ncbddd/ehdi/data.htm

Thompson, M., Thompson, G., & Vethivelu, S. (1989). A comparison of audiometric test methods for 2-year-old children. *Journal of Hearing and Speech Disorders, 54*, 174-179.

Trainor, L., & DesJardins, R. (2002). Pitch Characteristics of Infant-Directed Speech Affect Infants' Ability to Discriminate Vowels. *Psychonomic Bulletin & Review, 9*, 335-340.

Trehub, S. E., Bull, D., & Schneider, B. A. (1981). Infants' Detection of Speech in Noise. *Journal of Speech and Hearing Research, 24*, 202-206.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Development, 75*(4), 1067-1084.

Uhler, K. (2008a). (personal communication ed.). Boulder, CO.

Uhler, K. (2008b). *Longitudinal Study of Infant Speech Perception in Young Cochlear Implant Candidates: Three Case Studies.* Unpublished dissertation, University of Colorado, Boulder.

Werker, J., & Lelonde, C. (1988). Cross-language speech perception: initial capabilities and developmental change. *Developmental Psychology, 24*, 672-683.

Werker, J., & McLeod, P. J. (1989). Infant Preference for Both Male and Female Infant-Directed Talk: A Developmental Study of Attentional and Affective Responsiveness. *Canadian Journal of Psychology, 43*(2), 230-246.

Werker, J., Polka, L., & Pegg, J. (1997). The conditioned head turn procedure as a method for testing infant speech perception. *Early Development and Parenting, 6*, 171-178.

Werker, J., & Tees, R. (1984). Cross language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49-63.

Wickens, T. D. (2002). *Elementary Signal Detection Theory*. New York: Oxford University Press.

Zimmerman-Phillips, S., McConkey Robbins, A., & Observer, M. J. (2001, May). Infants and Implants: Listening Skills in Very Young Children. *Loud & Clear!, 4,* 1-8.

Zimmerman-Phillips, S., Osberger, M. I., & Robbins, A. M. (1997). Infant-Toddler Meaningful Auditory Integration Scale. Sylmar, CA: Advanced Bionics Corp.

Appendix A: Subject 1

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 93* | 0.93* | 0.89* | 2.98* | 8* | 8* |
| ba/pa | 43 | 0.53 | 0.06 | 0.15 | 30 | 30 |
| da/ga | 47 | 0.56* | 0.11 | 0.53 | 30 | 30 |

Table 1. Subject 1's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

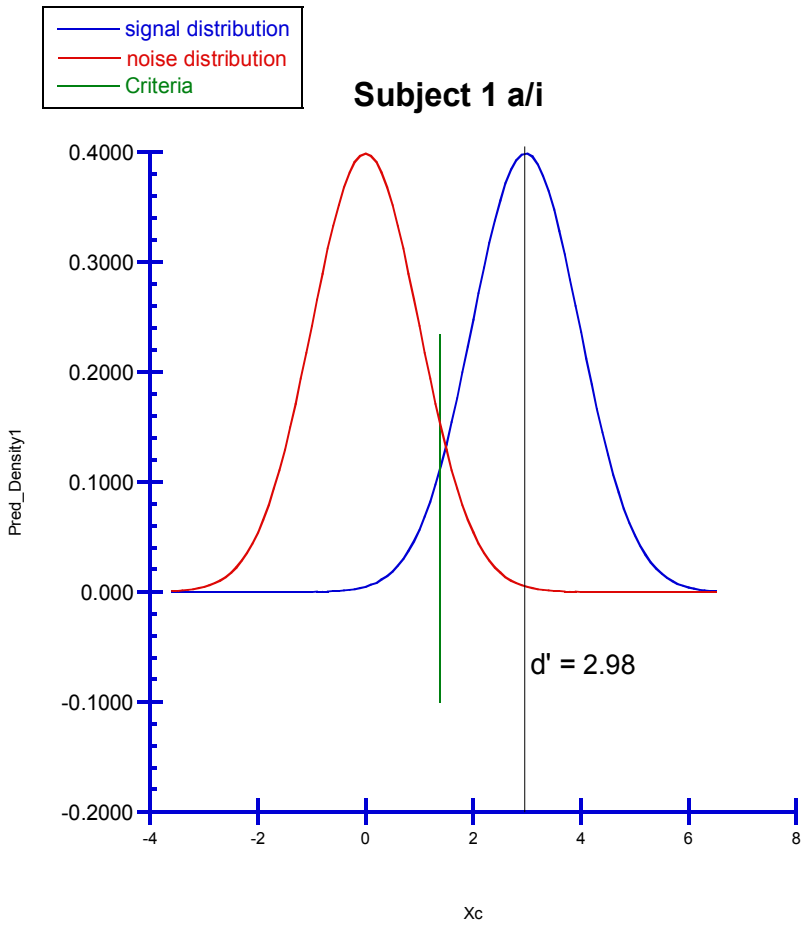


Figure 1. Subject 1's performance on the a/i contrast assessed via SDT.

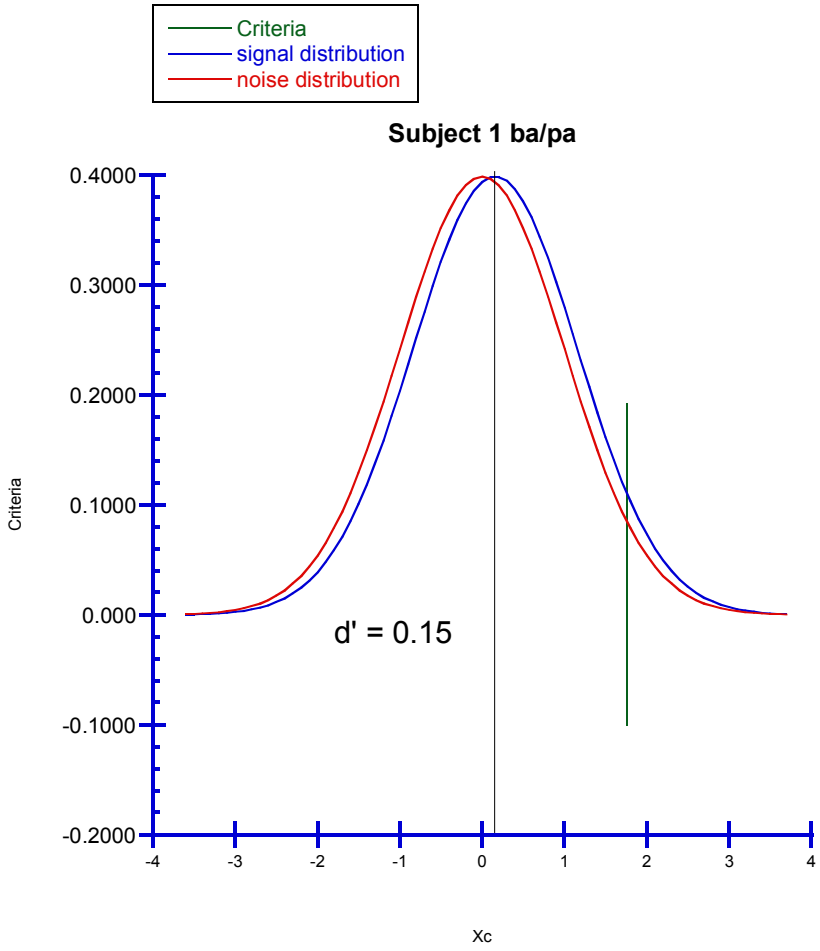


Figure 2. Subject 1's performance on the ba/pa contrast assessed via SDT.

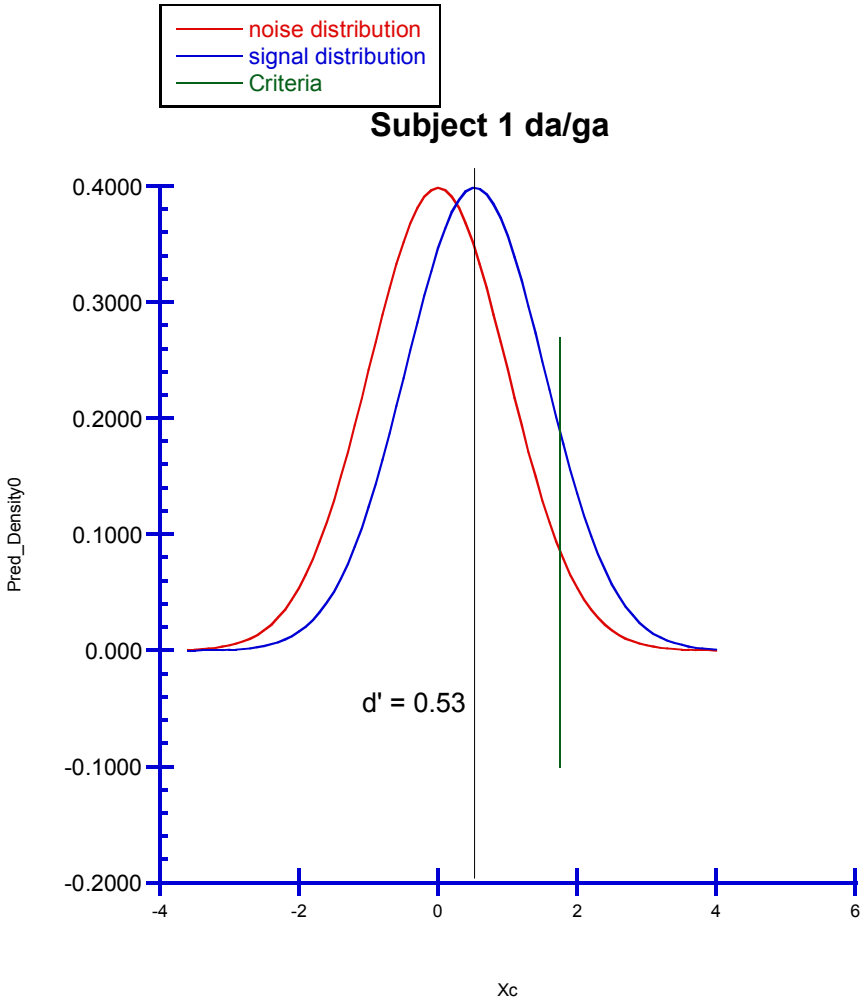


Figure 3. Subject 1's performance on the da/ga contrast assessed via SDT.

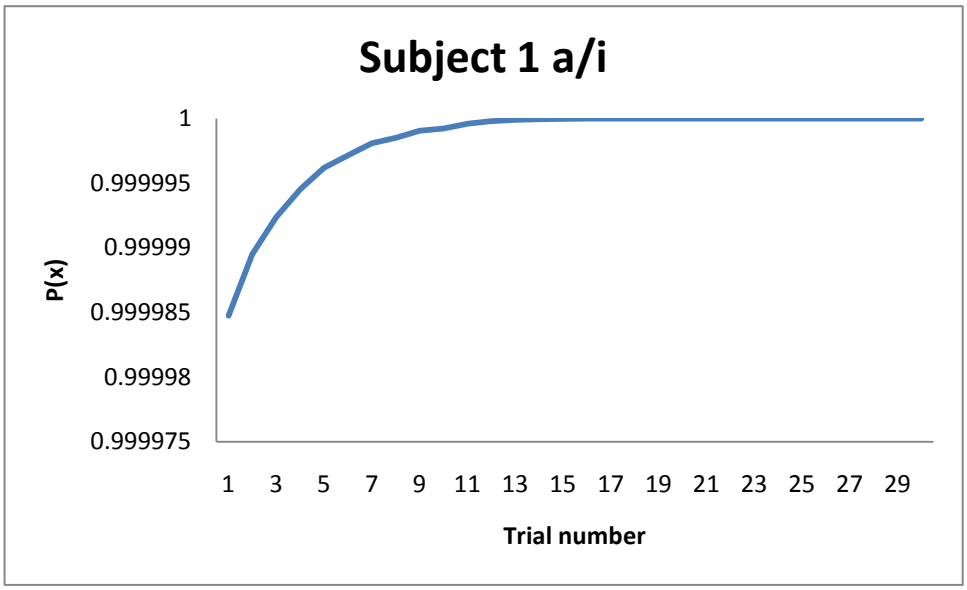


Figure 4. Cumulative distribution function for Subject 1's performance on the a/i contrast.

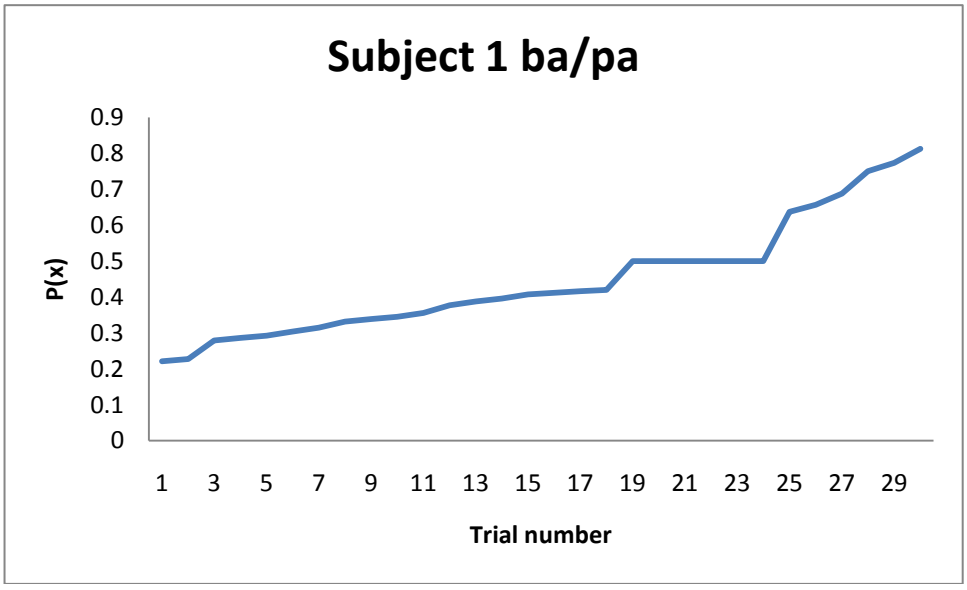


Figure 5. Cumulative distribution function for Subject 1's performance on the ba/pa contrast.

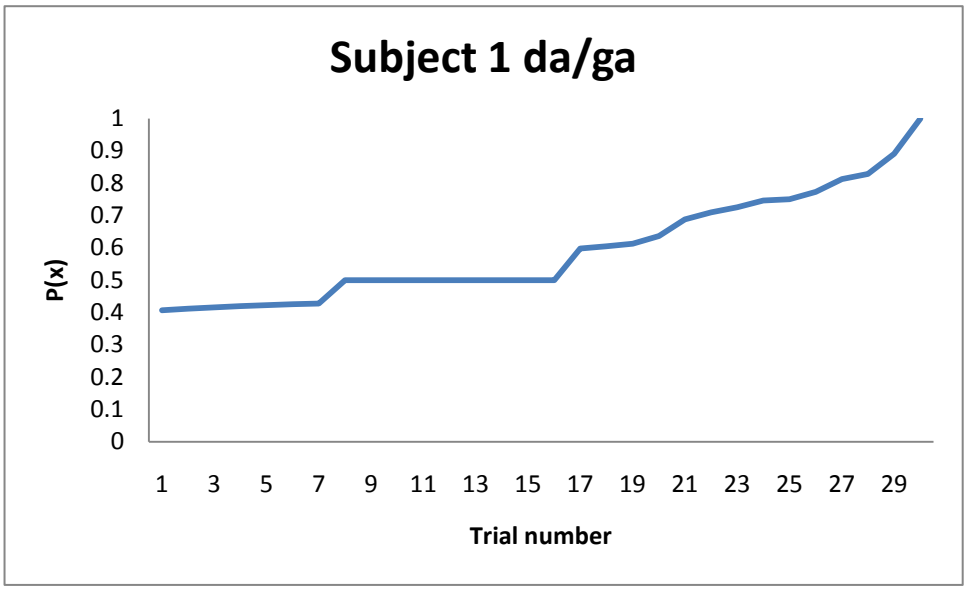


Figure 6. Cumulative distribution function for Subject 1's performance on the da/ga contrast.

Appendix B: Subject 2

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 93* | 0.89* | 0.83* | 2.56* | 8* | 8* |
| ba/pa | 37 | 0.46 | 0.06 | -0.54 | 30 | 30 |
| da/ga | 73* | 0.74* | 0.56* | 1.26* | 13* | 12* |

Table 1. Subject 2's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

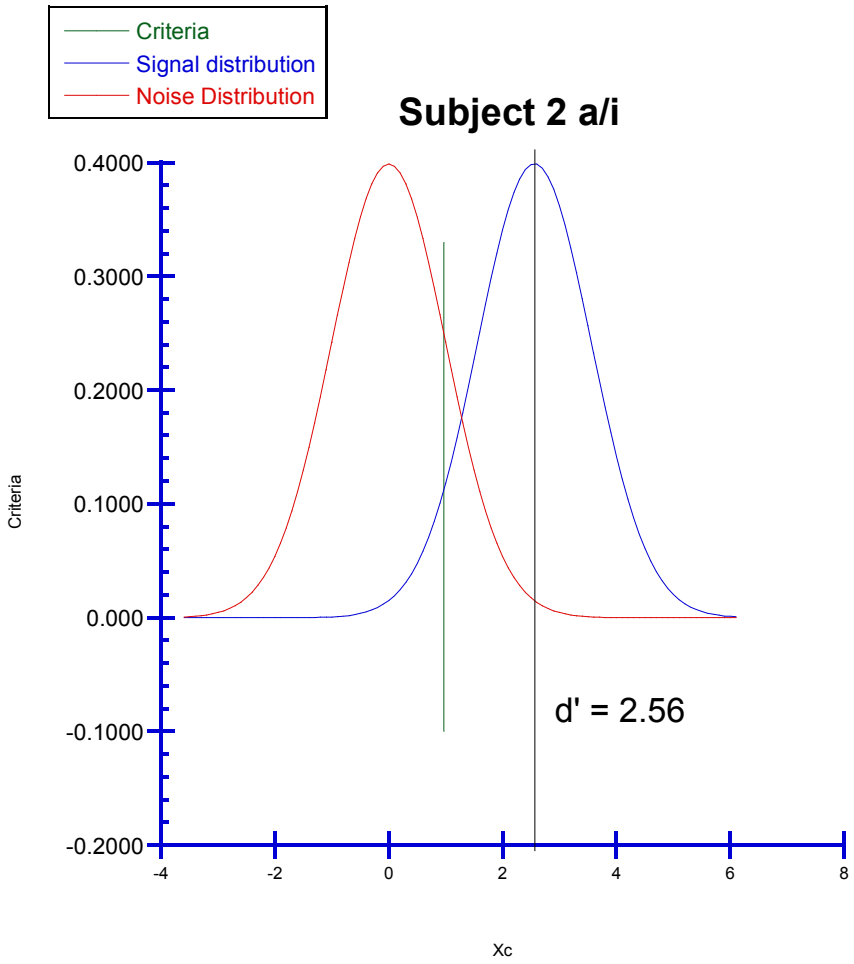


Figure 1. Subject 2's performance on the a/i contrast assessed via SDT.

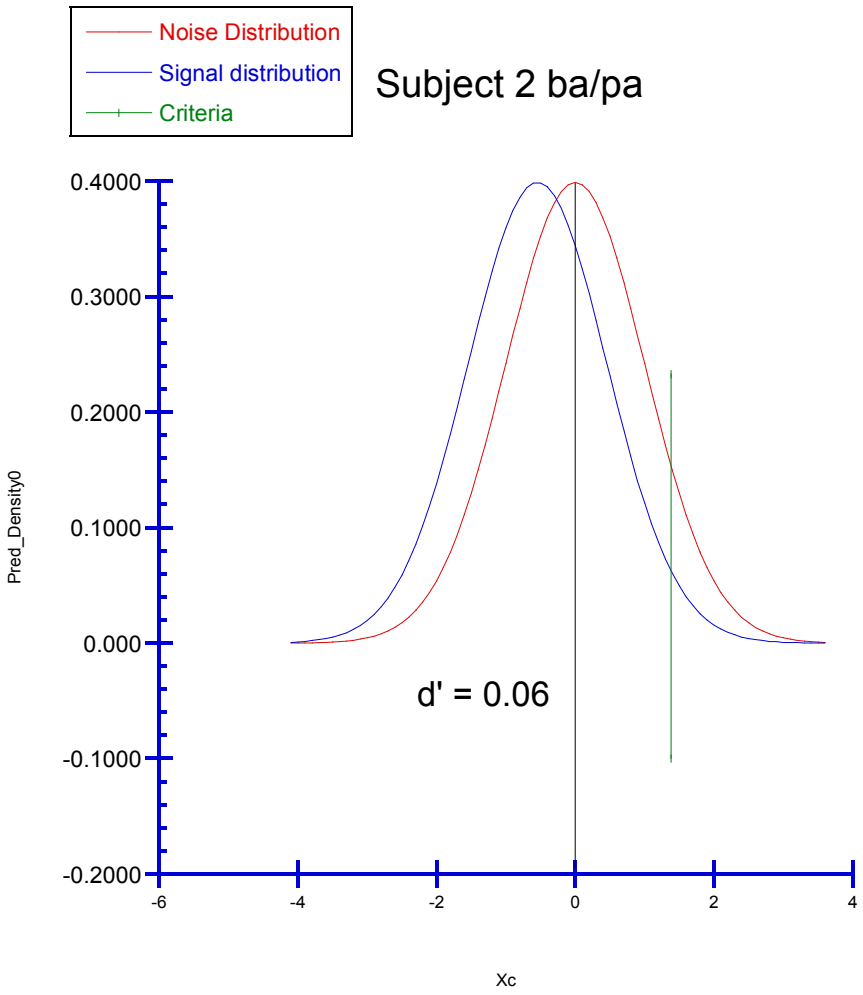


Figure 2. Subject 2's performance on the ba/pa contrast assessed via SDT.

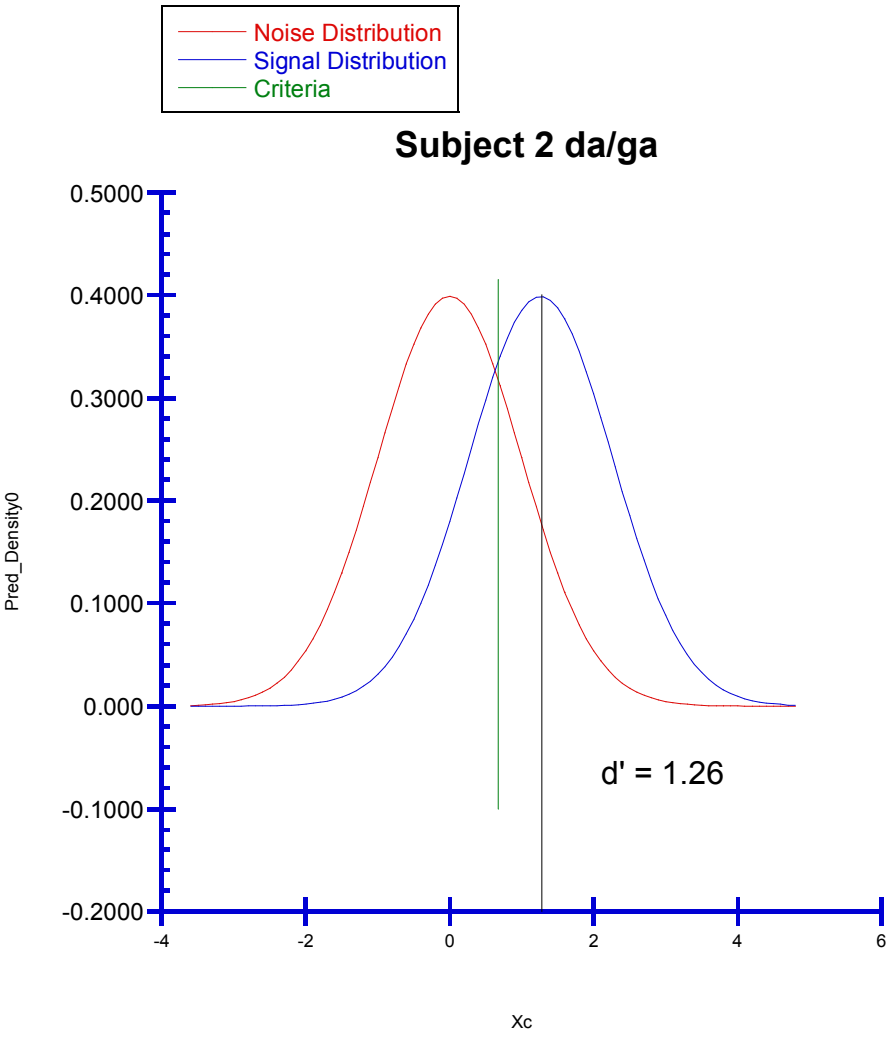


Figure 3. Subject 2's performance on the da/ga contrast assessed via SDT.

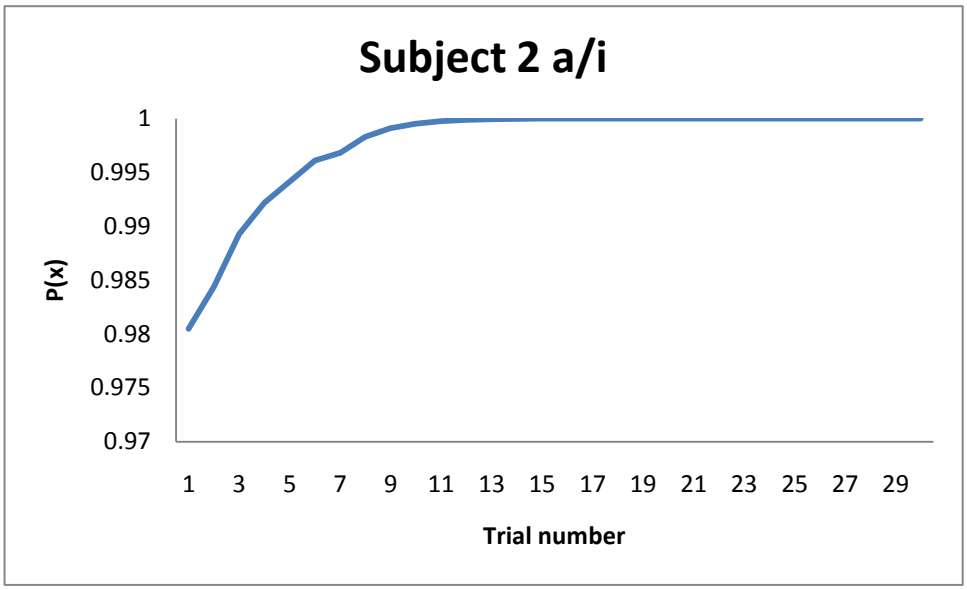


Figure 4. Cumulative distribution function for Subject 2's performance on the a/i contrast.

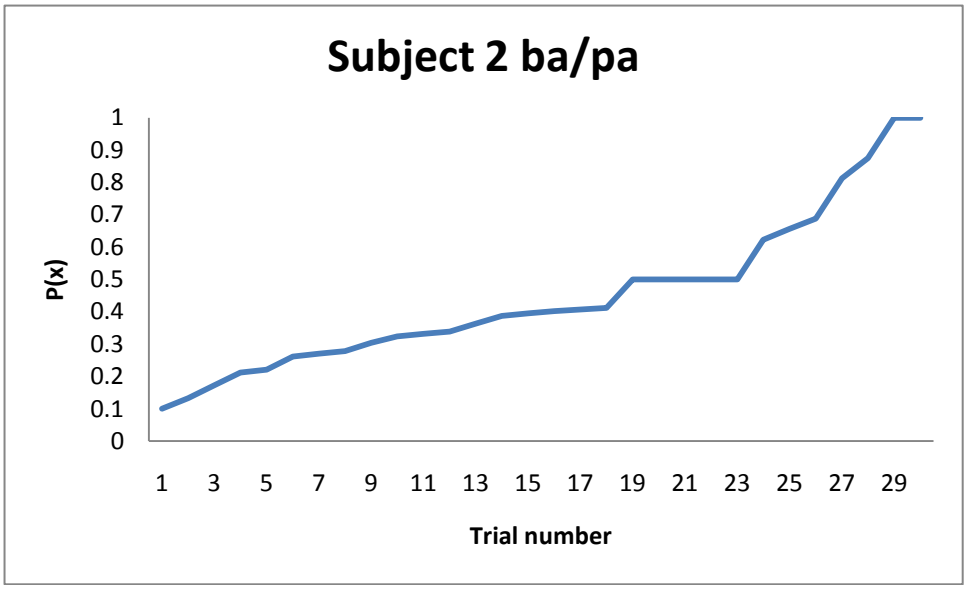


Figure 5. Cumulative distribution function for Subject 2's performance on the ba/pa contrast.

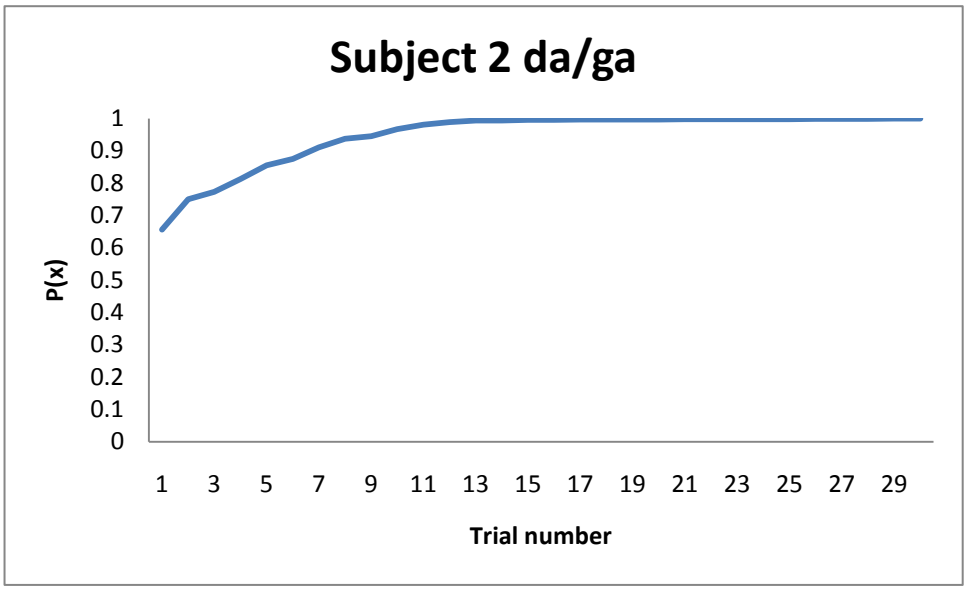


Figure 6. Cumulative distribution function for Subject 2's performance on the da/ga contrast.

Appendix C: Subject 3

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 67* | 0.71* | 0.44* | 1.38* | 8* | 8* |
| ba/pa | 87* | 0.84* | 0.78* | 0.71* | 8* | 8* |
| da/ga | 40 | 0.3 | 0 | -0.33 | 30 | 30 |

Table 1. Subject 3's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

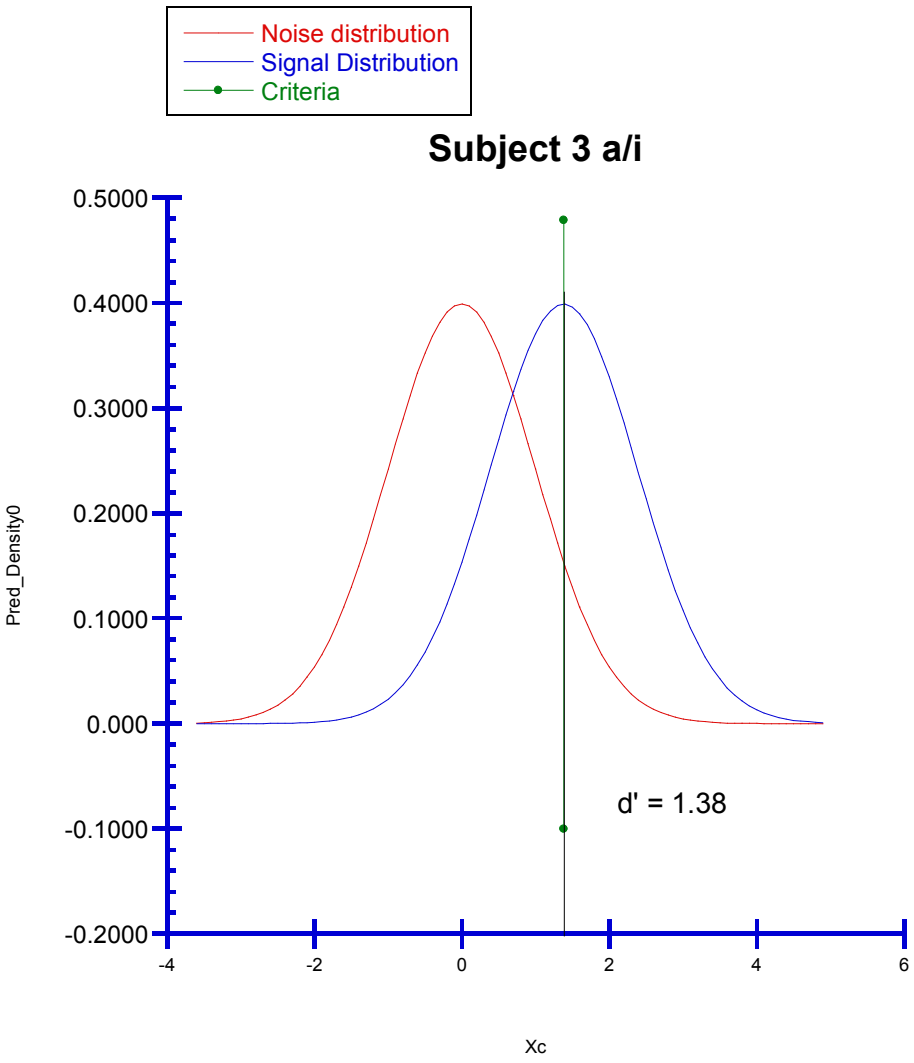


Figure 1. Subject 3's performance on the a/i contrast assessed via SDT.

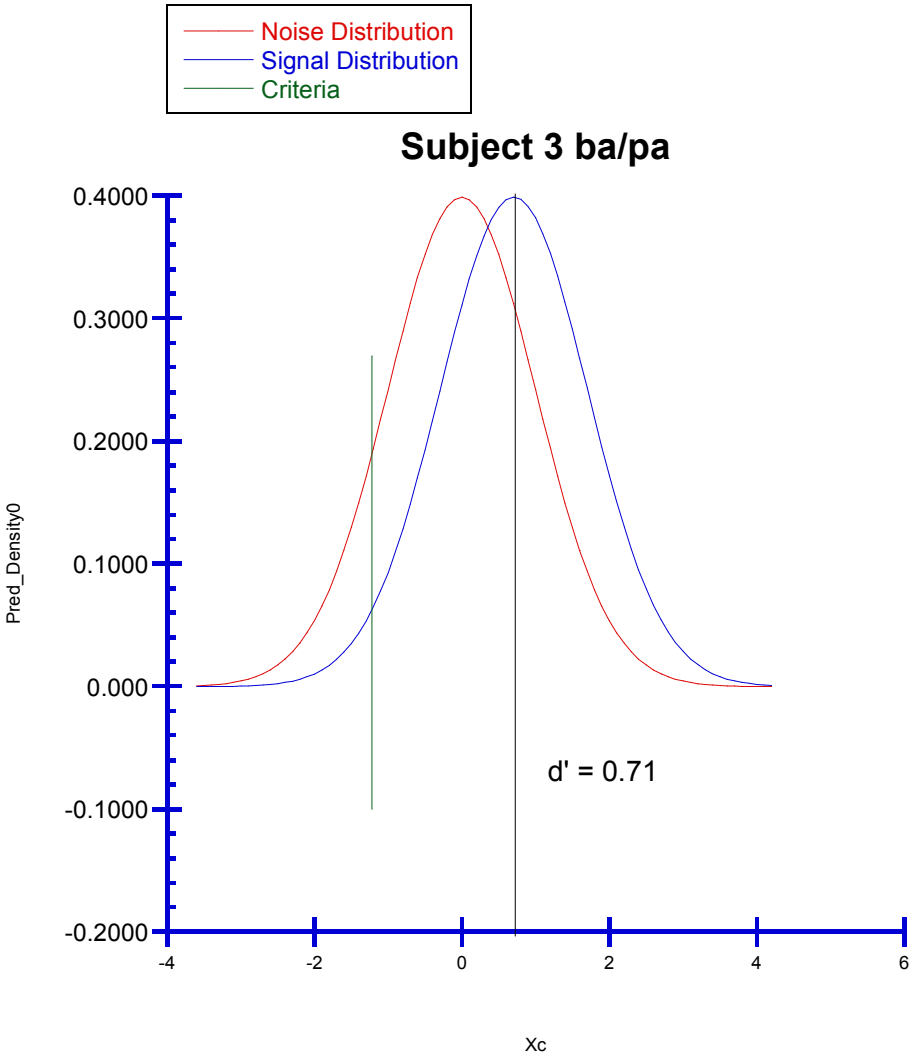


Figure 2. Subject 3's performance on the ba/pa contrast assessed via SDT.

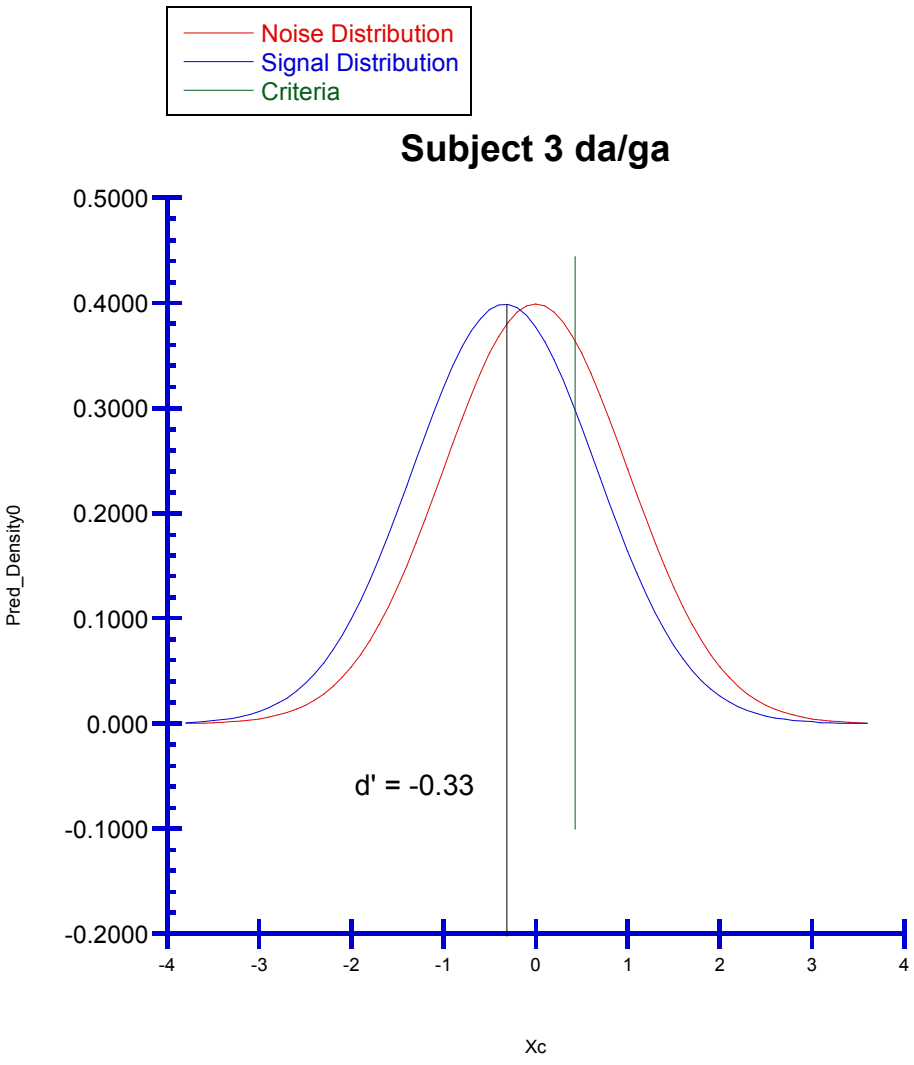


Figure 3. Subject 3's performance on the da/ga contrast assessed via SDT.

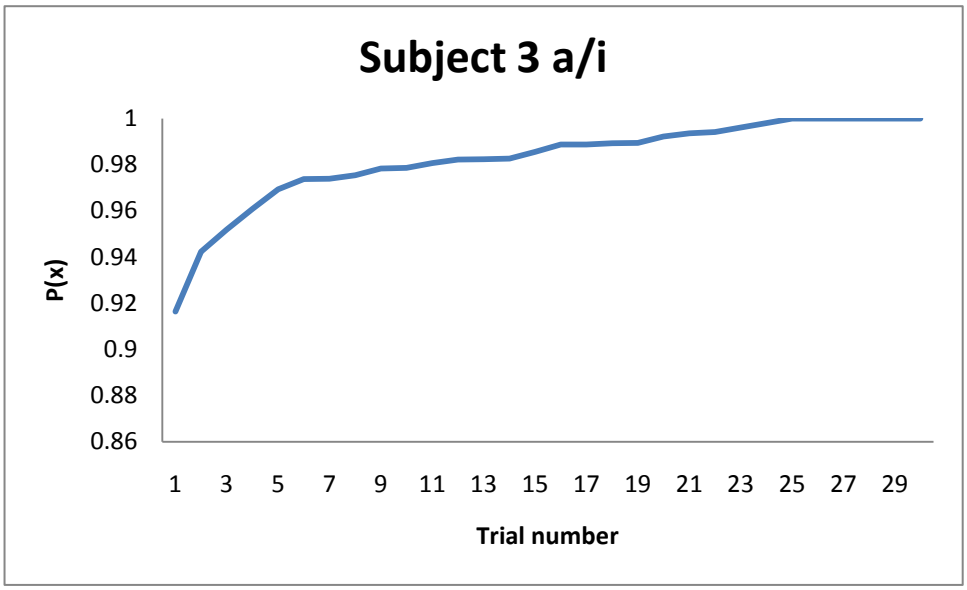


Figure 4. Cumulative distribution function for Subject 3's performance on the a/i contrast.

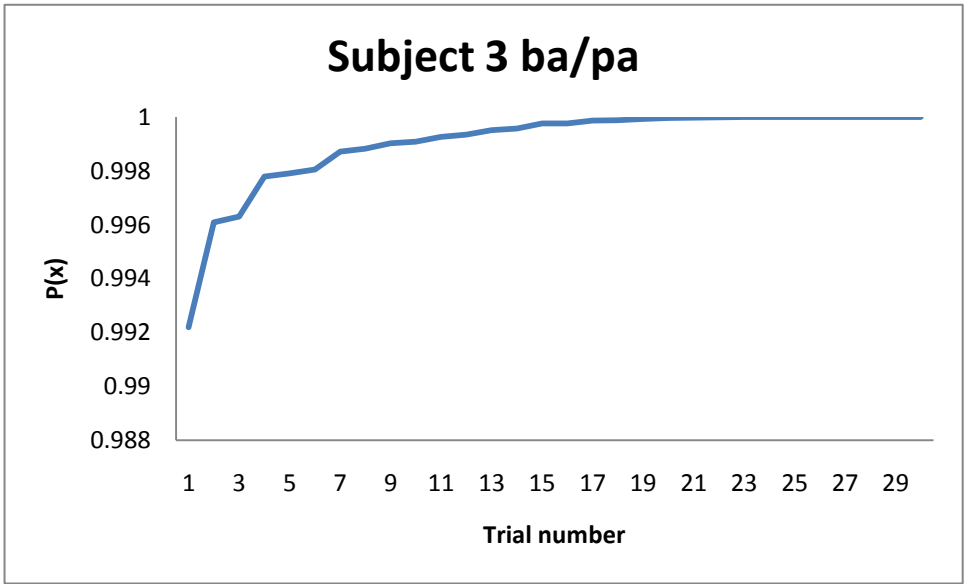


Figure 5. Cumulative distribution function for Subject 3's performance on the ba/pa contrast.

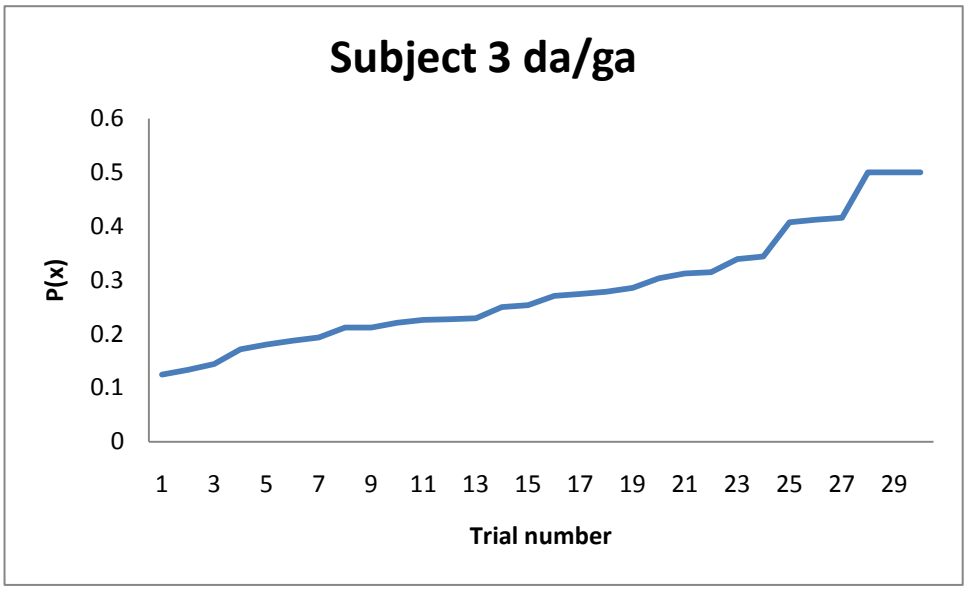


Figure 6. Cumulative distribution function for Subject 3's performance on the da/ga contrast.

Appendix D: Subject 4

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 83* | 0.82* | 0.72* | 1.89* | 8* | 8* |
| ba/pa | 47 | 0.52 | 0.11 | 0.09 | 30 | 30 |
| da/ga | 47 | 0.55 | 0.11 | 0.42 | 30 | 8* |

Table 1. Subject 4's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

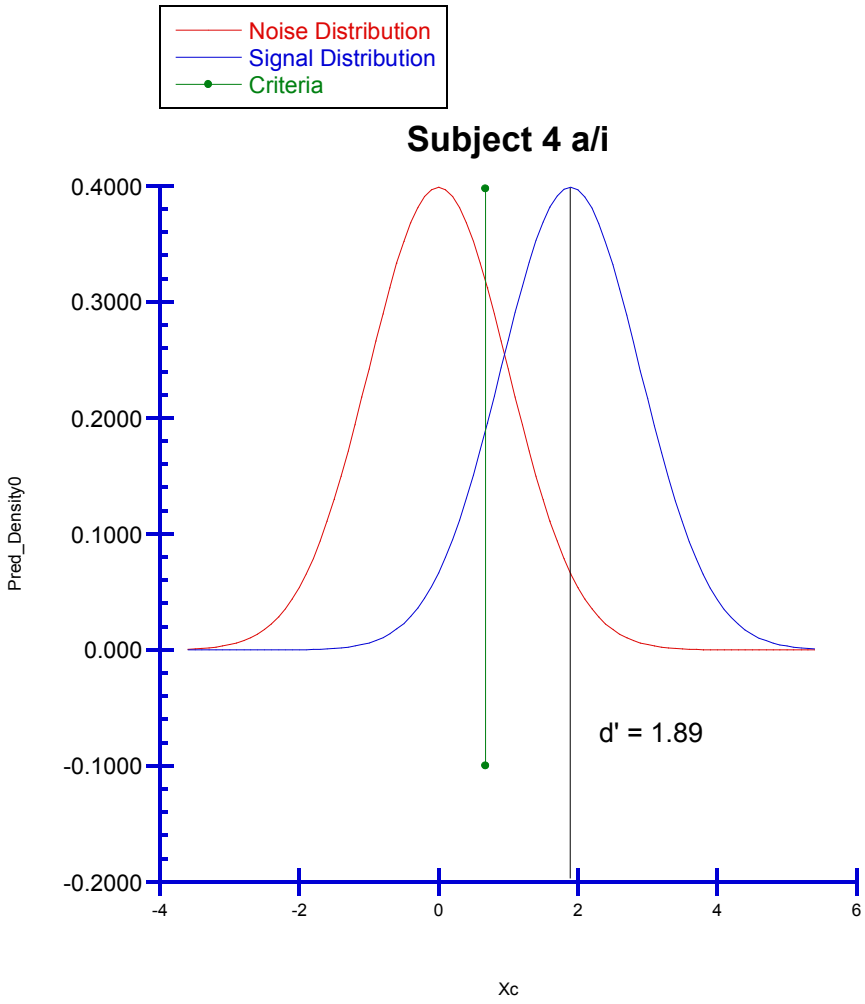


Figure 1. Subject 4's performance on the a/i contrast assessed via SDT.

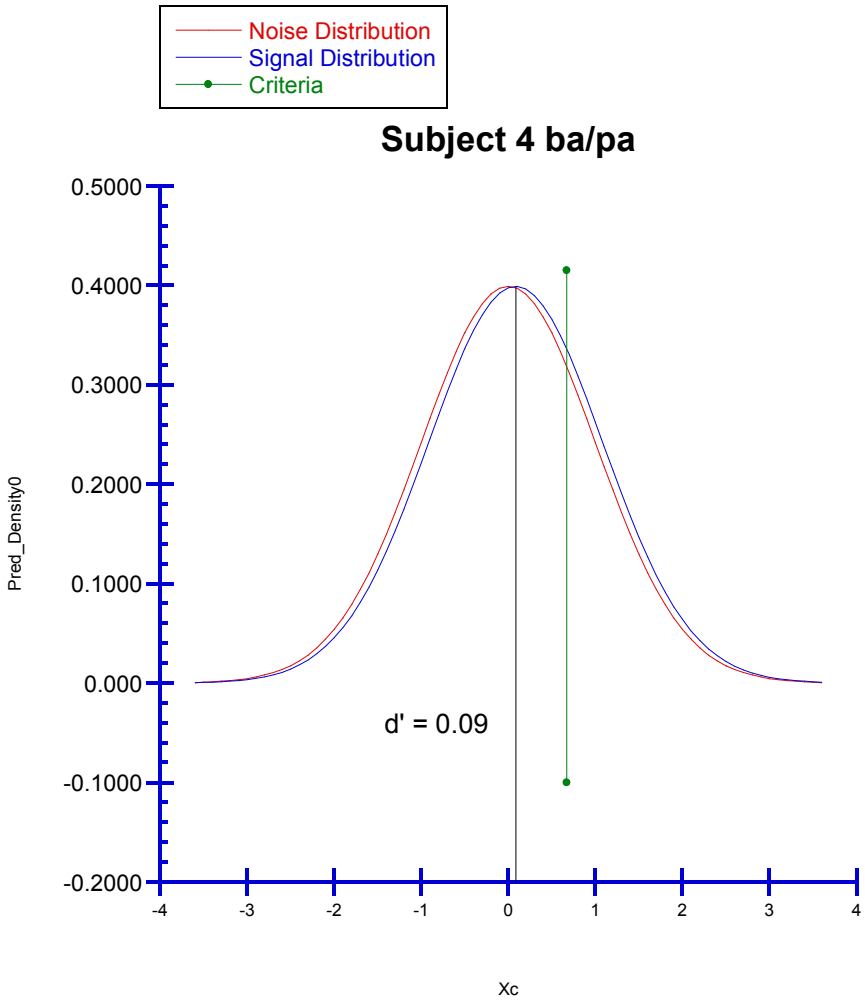


Figure 2. Subject 4's performance on the ba/pa contrast assessed via SDT.

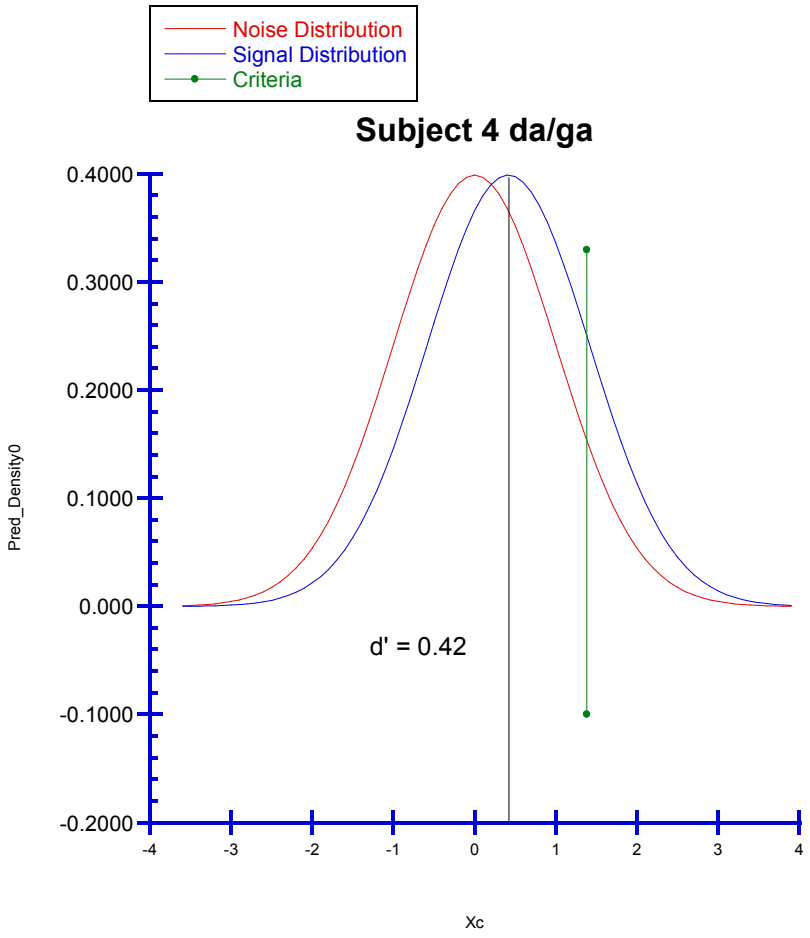


Figure 3. Subject 4's performance on the da/ga contrast assessed via SDT.

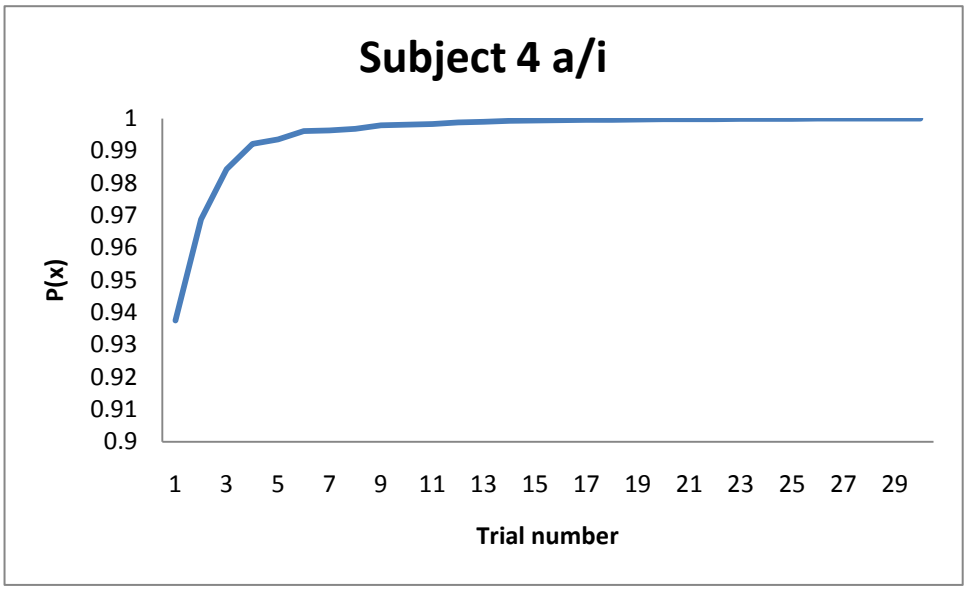


Figure 4. Cumulative distribution function for Subject 4's performance on the a/i contrast.

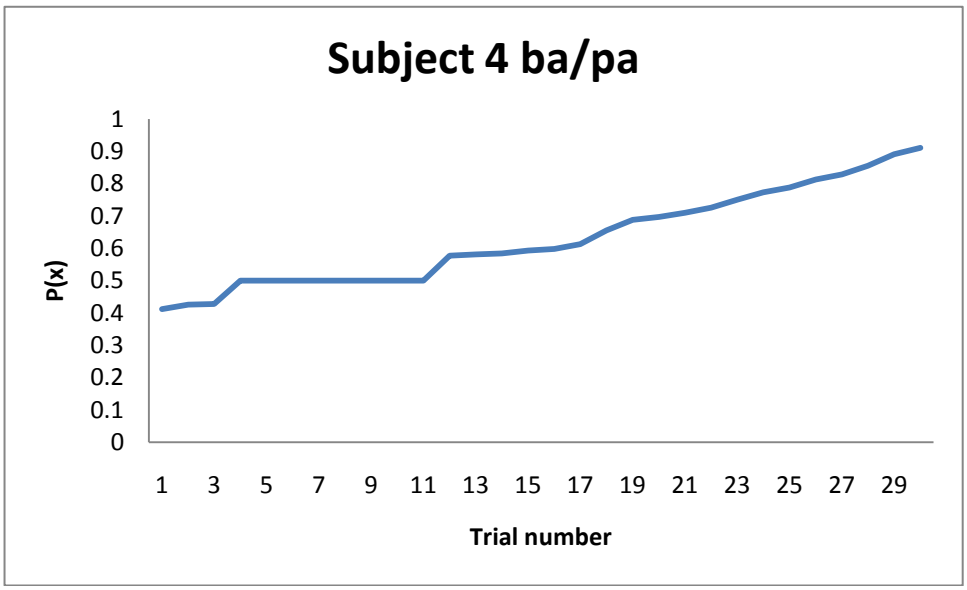


Figure 5. Cumulative distribution function for Subject 4's performance on the ba/pa contrast.

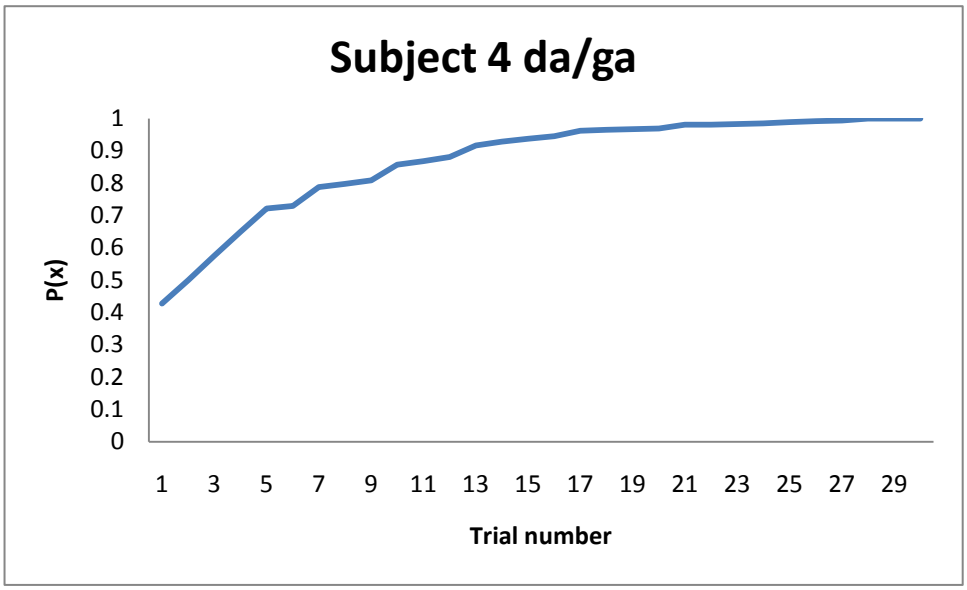


Figure 6. Cumulative distribution function for Subject 4's performance on the da/ga contrast.

Appendix E: Subject 5

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 90* | 0.89* | 0.83* | 2.56* | 10* | 12* |
| ba/pa | 83* | 0.81* | 0.72* | 2.5* | 8* | 8* |
| da/ga | 83* | 0.85* | 0.72* | 2.15* | 13* | 12* |

Table 1. Subject 5's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

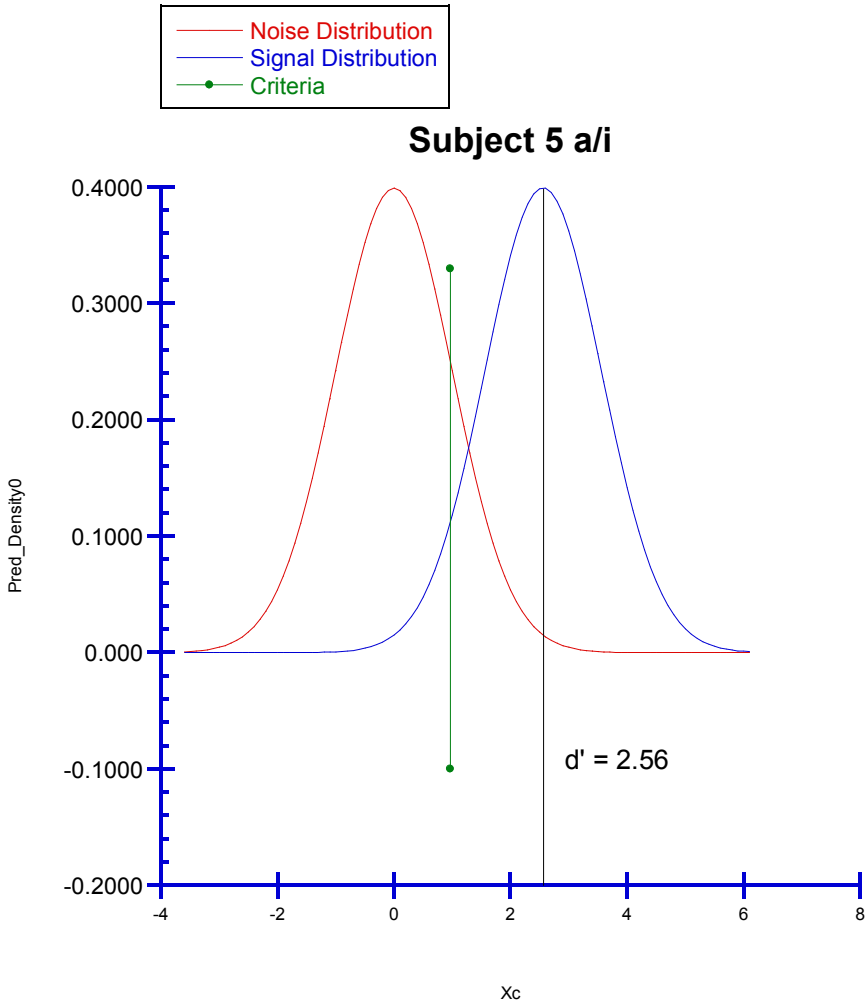


Figure 1. Subject 5's performance on the a/i contrast assessed via SDT.

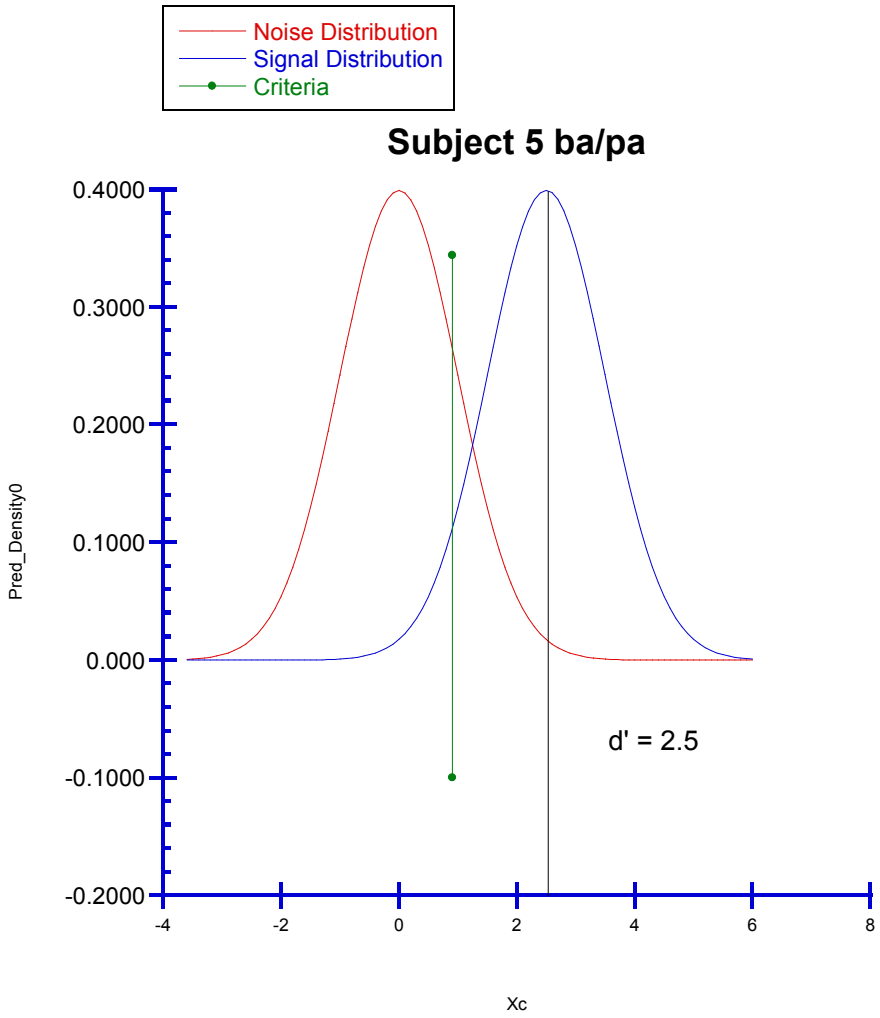


Figure 2. Subject 5's performance on the ba/pa contrast assessed via SDT.

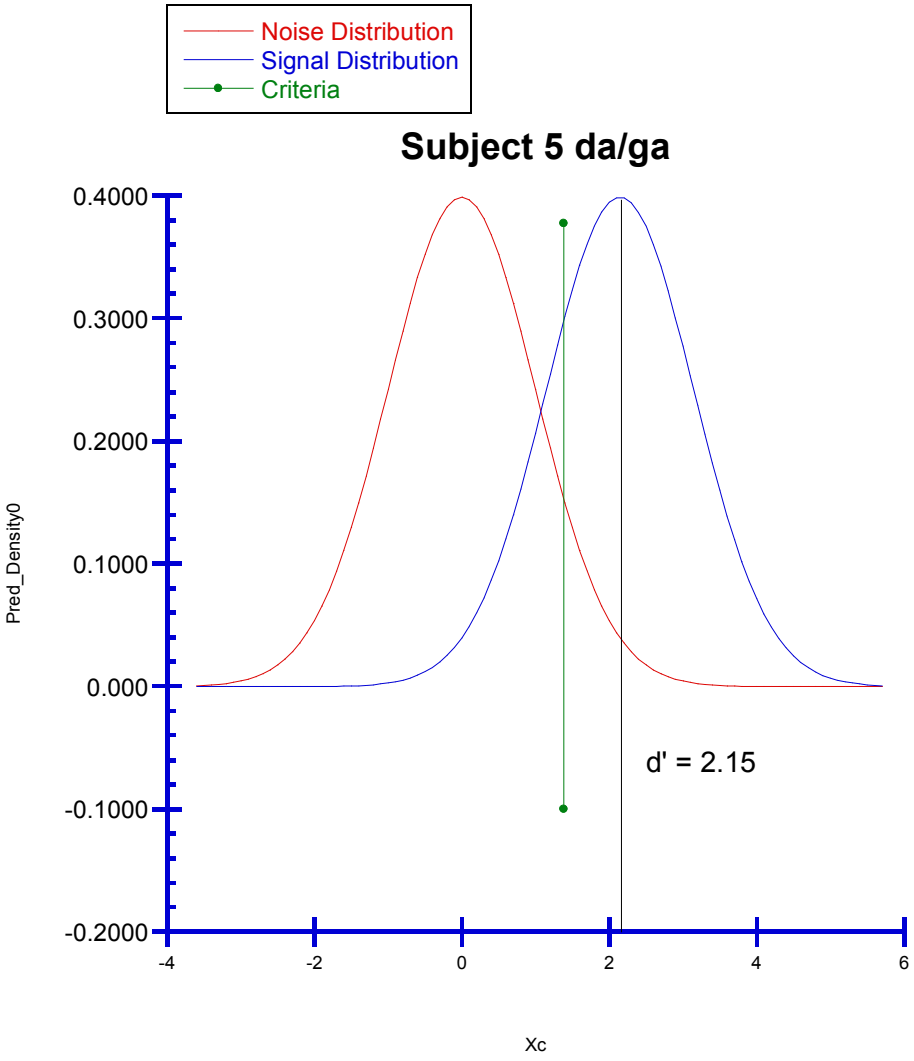


Figure 3. Subject 5's performance on the da/ga contrast assessed via SDT.

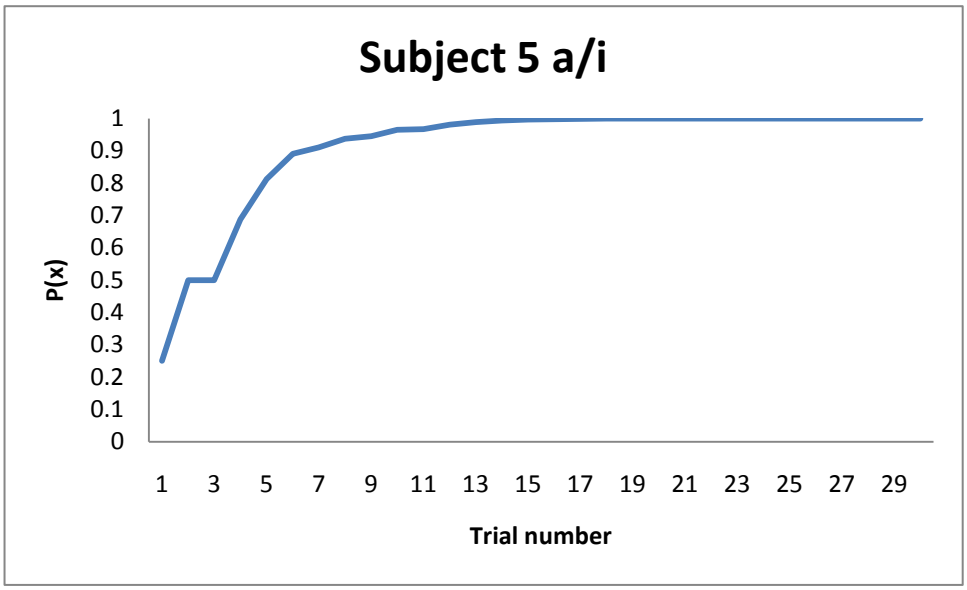


Figure 4. Cumulative distribution function for Subject 5's performance on the a/i contrast.

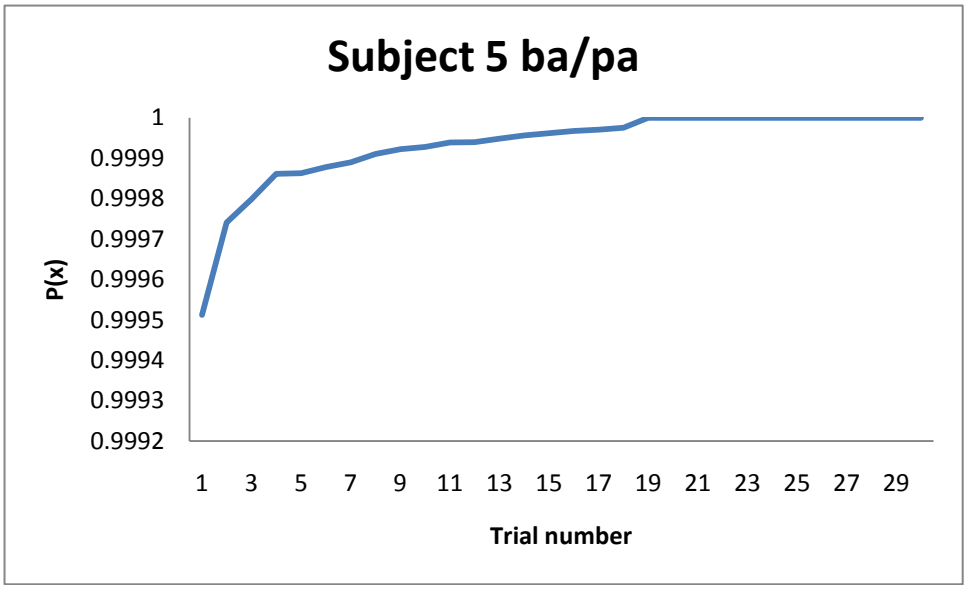


Figure 5. Cumulative distribution function for Subject 5's performance on the ba/pa contrast.

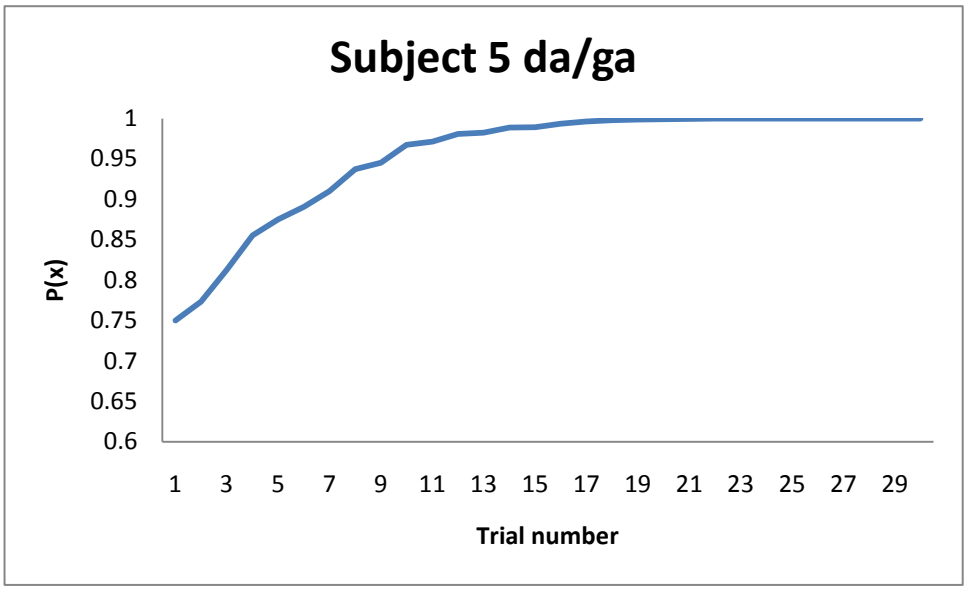


Figure 6. Cumulative distribution function for Subject 5's performance on the da/ga contrast.

Appendix F: Subject 6

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 70* | 0.71* | 0.5* | 1.1* | 30 | 8* |
| ba/pa | 67* | 0.68* | 0.44 | 0.96* | 30 | 8* |
| da/ga | 57 | 0.61* | 0.28 | 0.69 | 30 | 12* |

Table 1. Subject 6's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

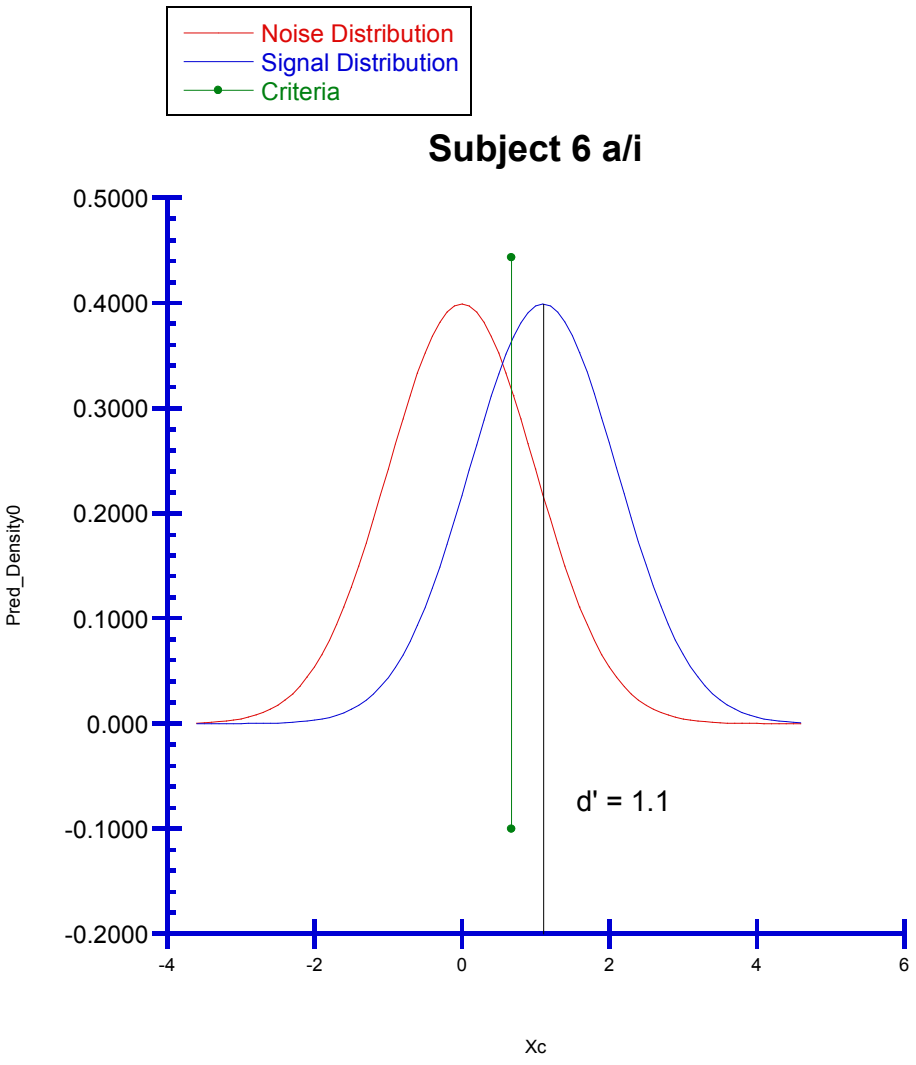


Figure 1. Subject 6's performance on the a/i contrast assessed via SDT.

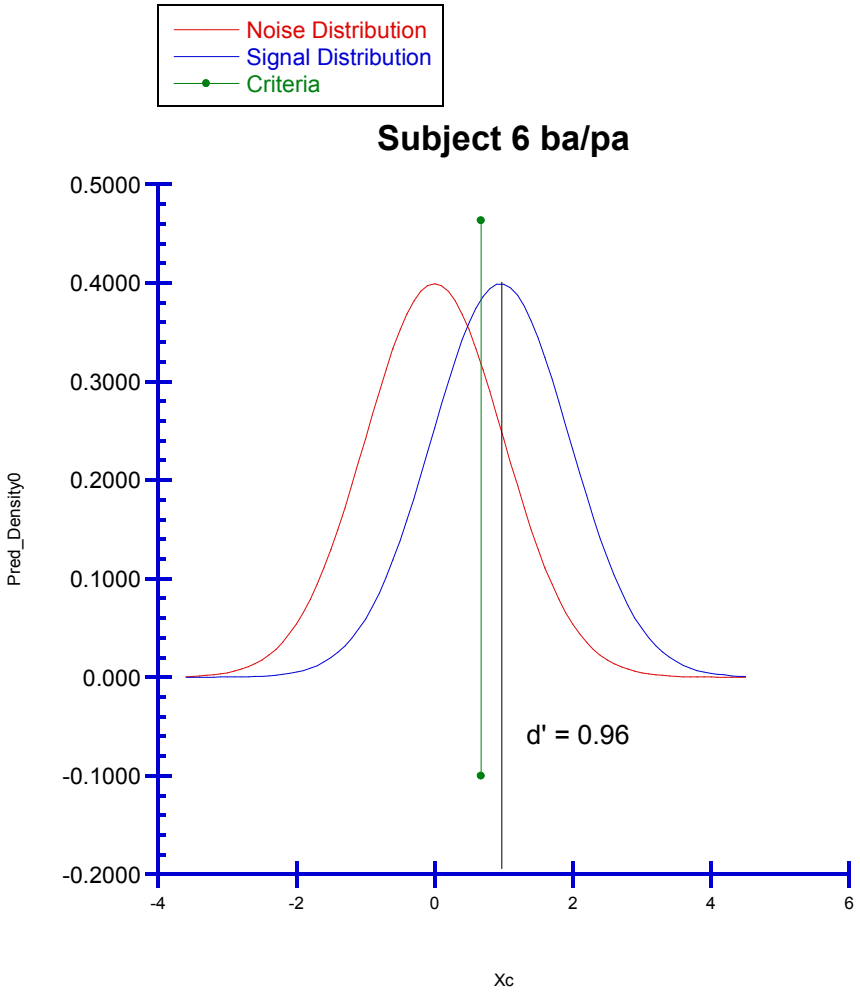


Figure 2. Subject 6's performance on the ba/pa contrast assessed via SDT.

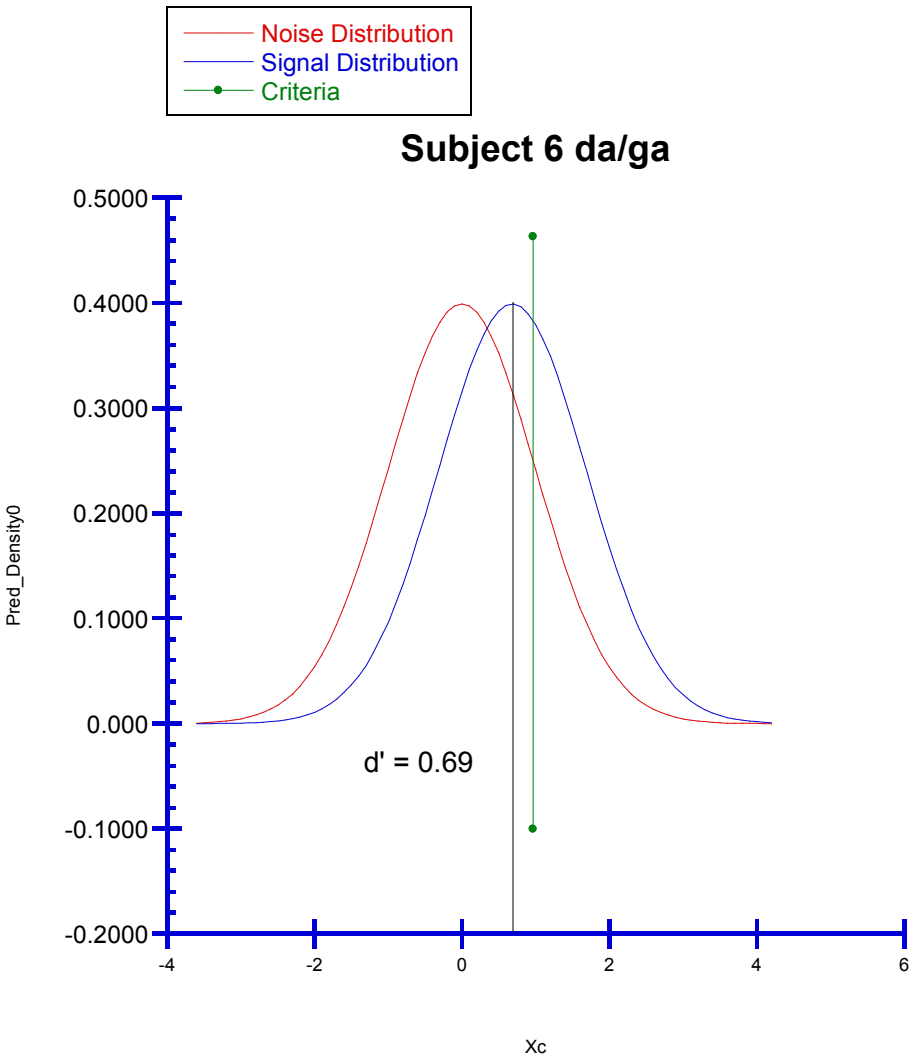


Figure 3. Subject 6's performance on the da/ga contrast assessed via SDT.

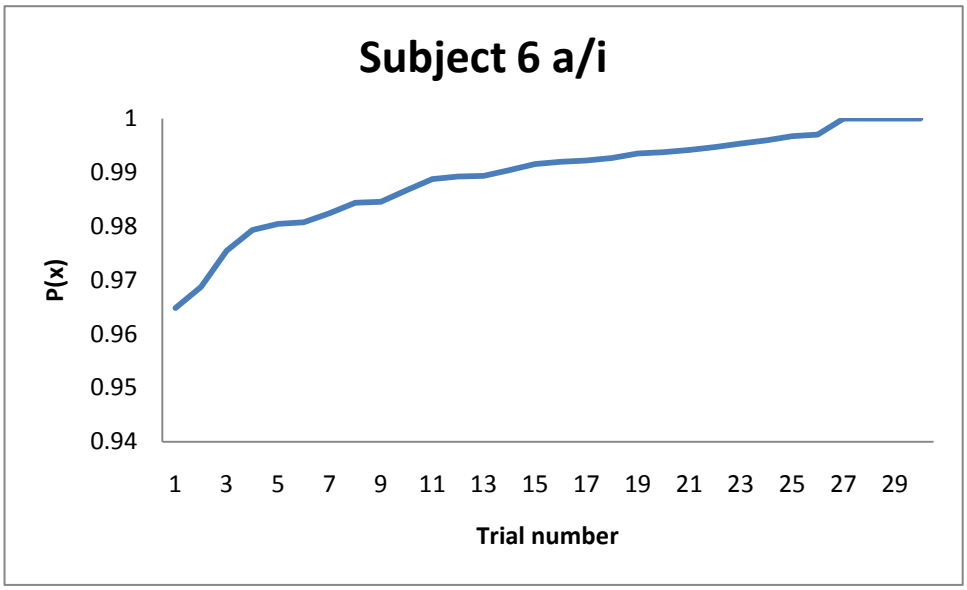


Figure 4. Cumulative distribution function for Subject 6's performance on the a/i contrast.

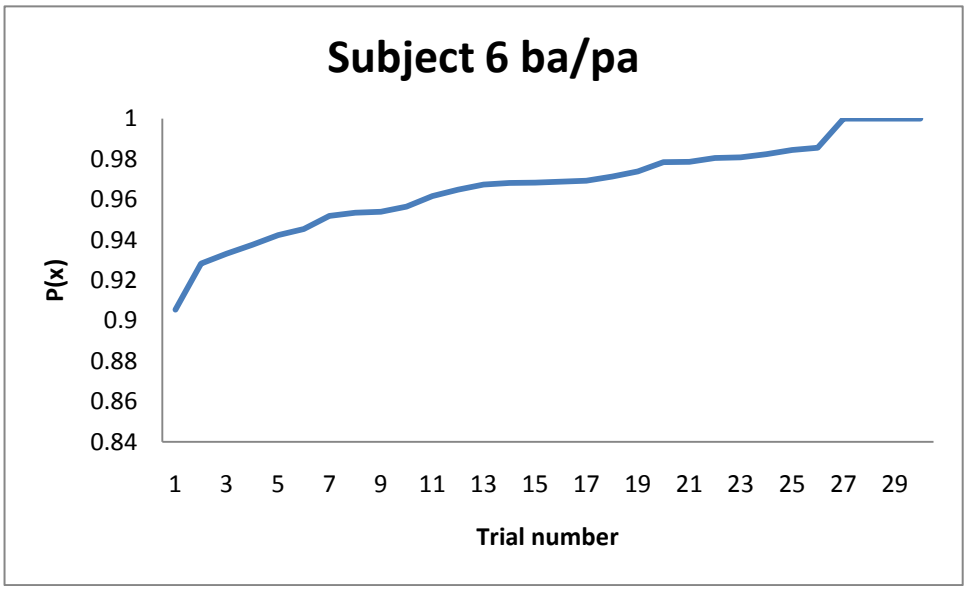


Figure 5. Cumulative distribution function for Subject 6's performance on the ba/pa contrast.

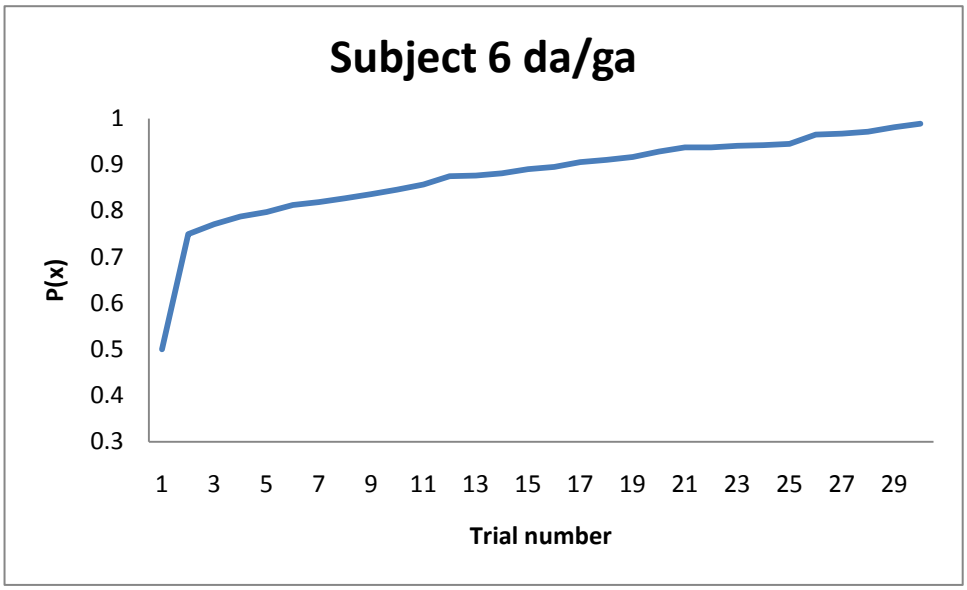


Figure 6. Cumulative distribution function for Subject 6's performance on the da/ga contrast.

Appendix G: Subject 7

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 83* | 0.79* | 0.72* | 2.14* | 8* | 8* |
| ba/pa | 50 | 0.57* | 0.17 | 0.62 | 30 | 30 |
| da/ga | 36 | 0.42 | -0.06 | -0.53 | 30 | 30 |

Table 1. Subject 7's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

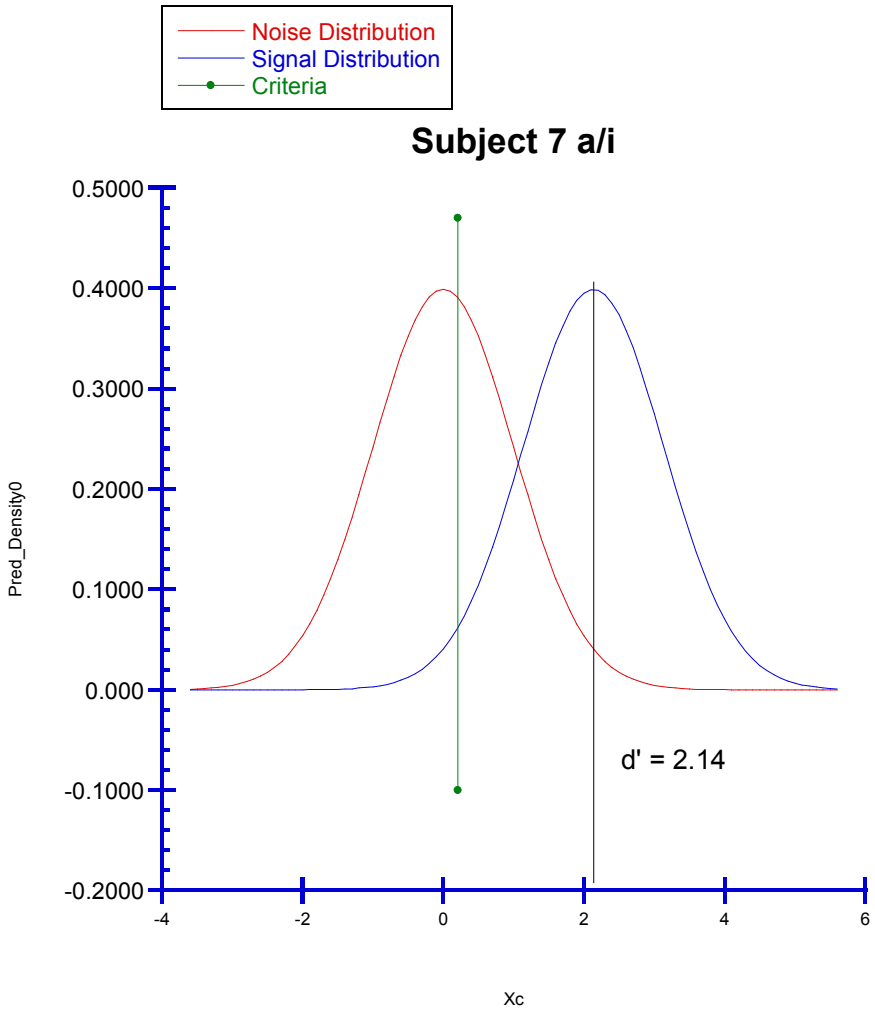


Figure 1. Subject 7's performance on the a/i contrast assessed via SDT.

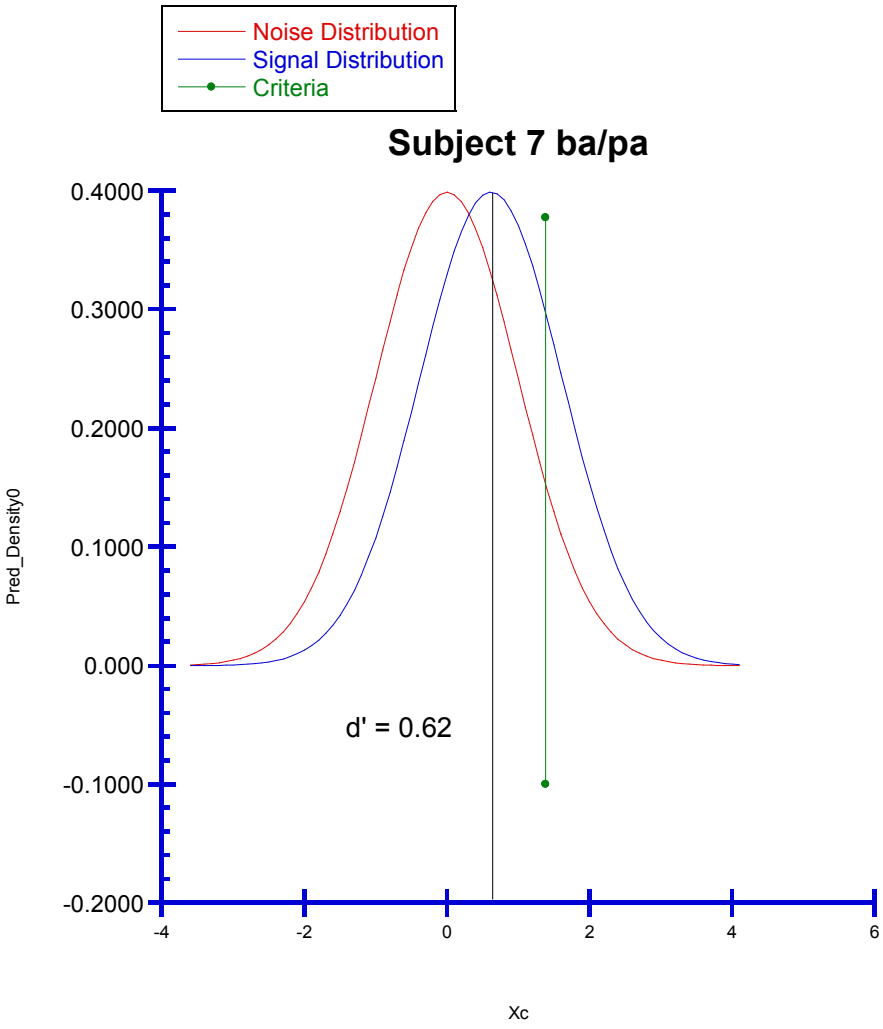


Figure 2. Subject 7's performance on the ba/pa contrast assessed via SDT.

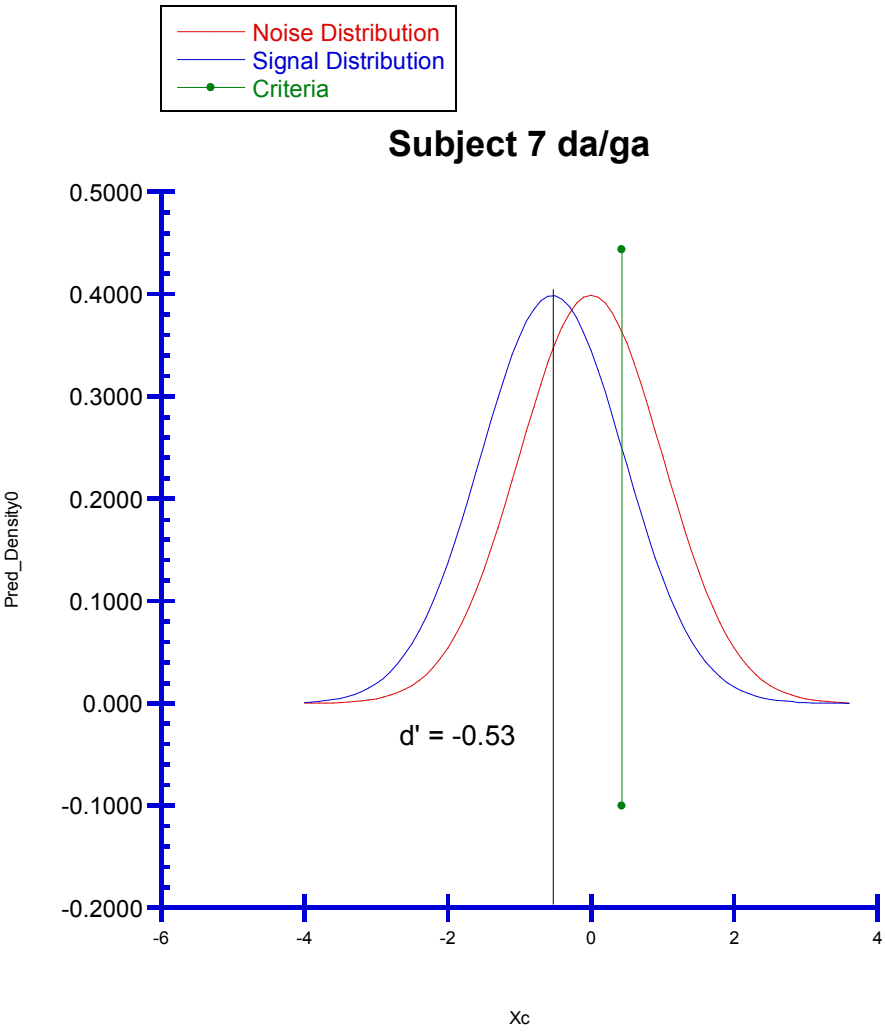


Figure 3. Subject 7's performance on the da/ga contrast assessed via SDT.

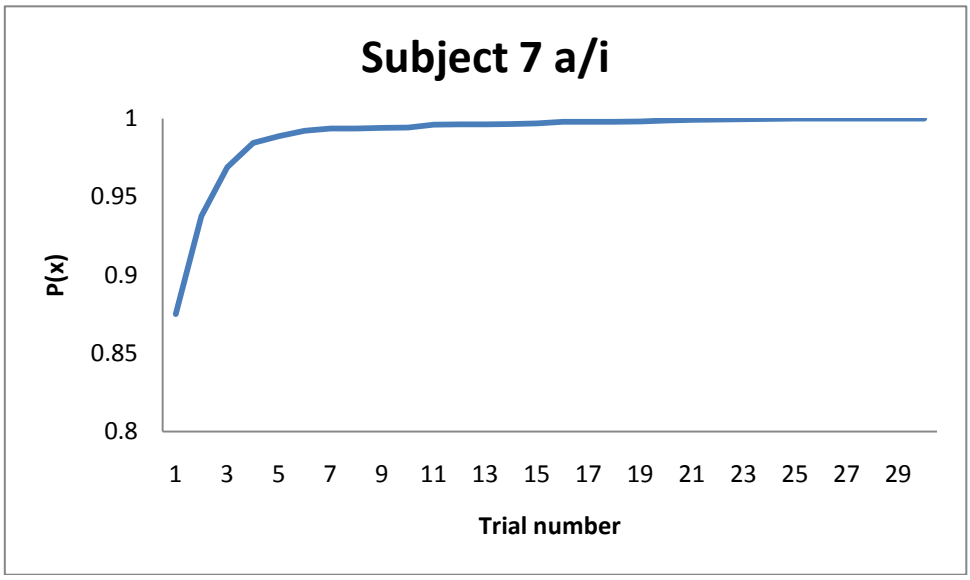


Figure 4. Cumulative distribution function for Subject 7's performance on the a/i contrast.

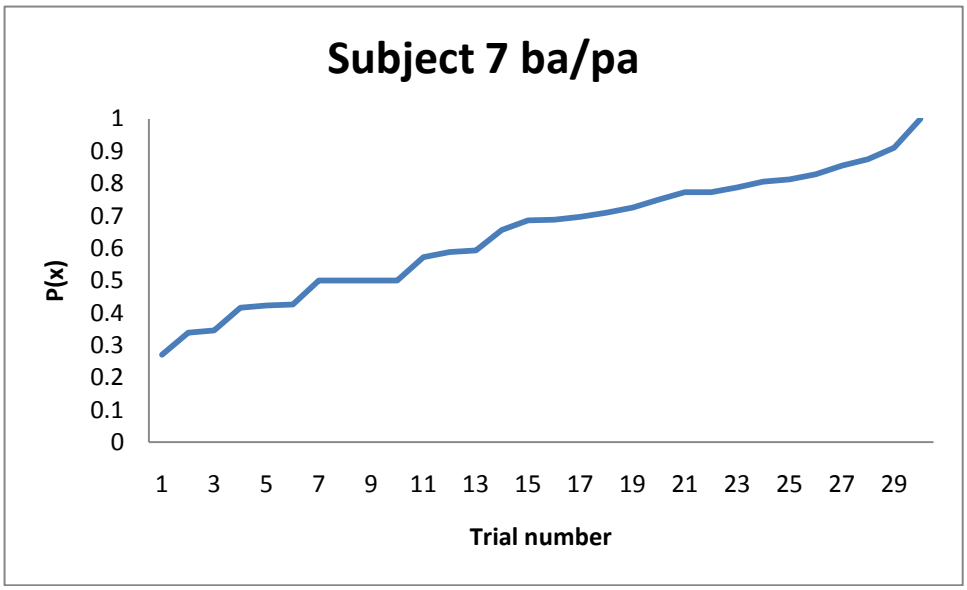


Figure 5. Cumulative distribution function for Subject 7's performance on the ba/pa contrast.

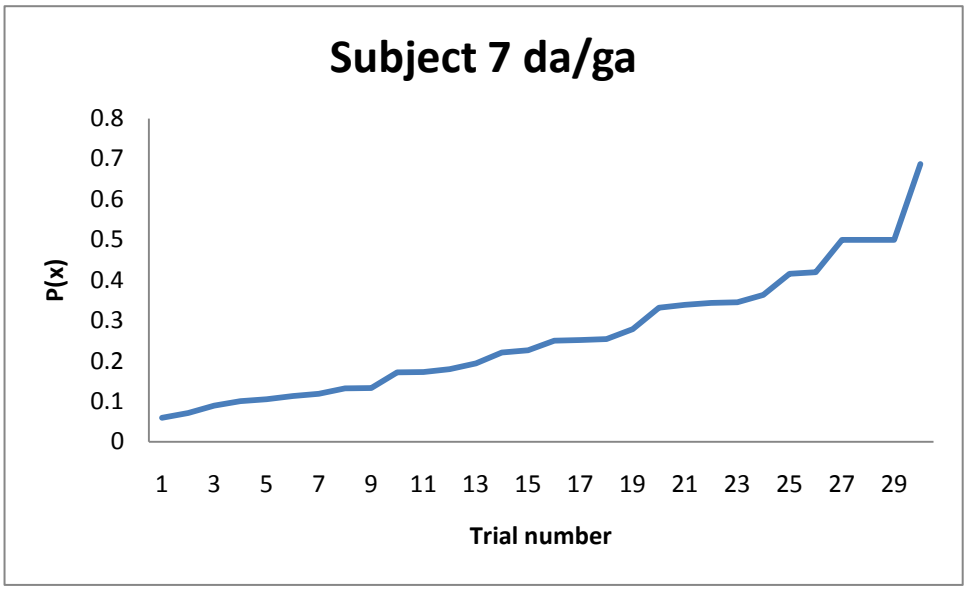


Figure 6. Cumulative distribution function for Subject 7's performance on the da/ga contrast.

Appendix H: Subject 8

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 83* | 0.82* | 0.72* | 1.9* | 8* | 8* |
| ba/pa | 73* | 0.74* | 0.56 | 1.26* | 30 | 8* |
| da/ga | 63* | 0.64* | 0.39 | 0.71* | 19* | 19* |

Table 1. Subject 8's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

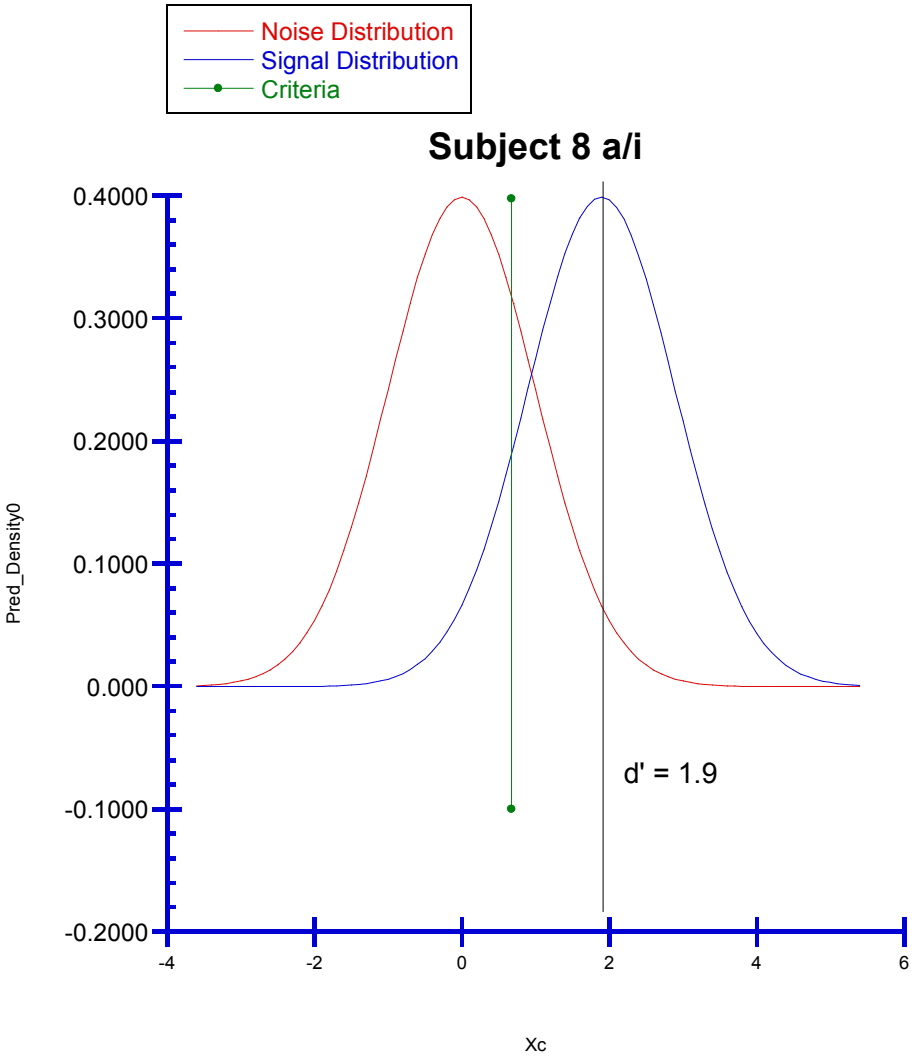


Figure 1. Subject 8's performance on the a/i contrast assessed via SDT.

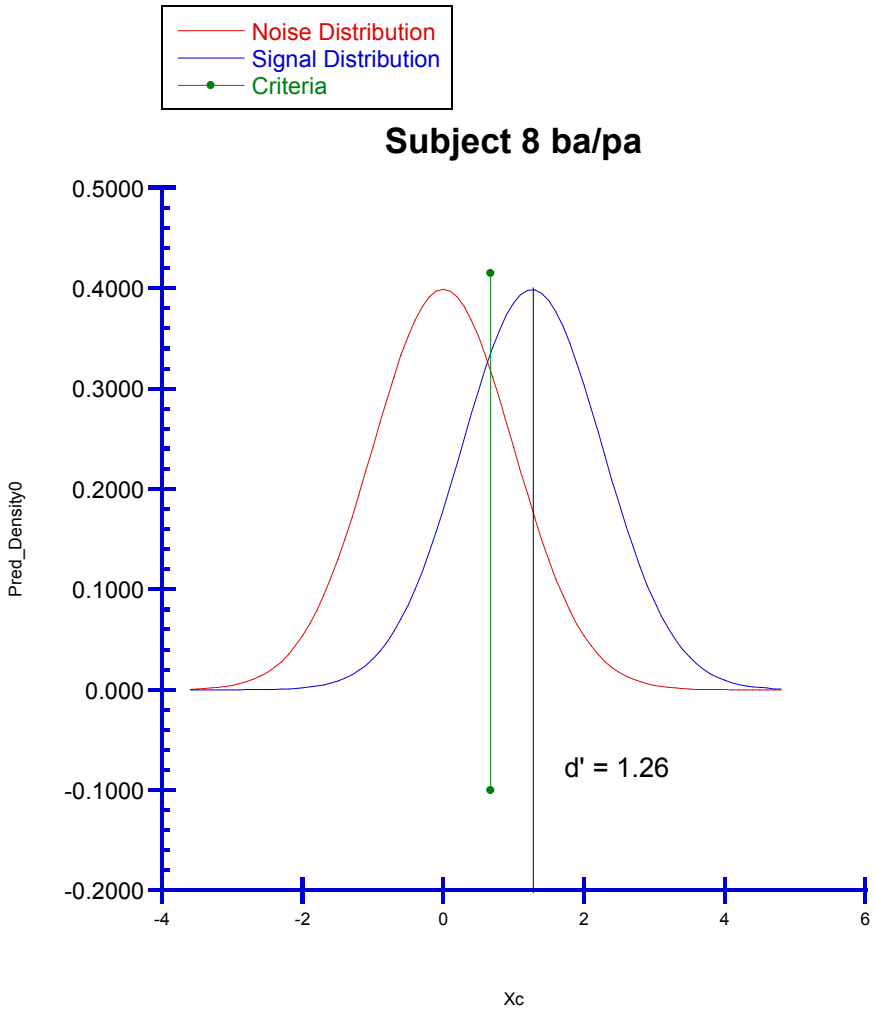


Figure 2. Subject 8's performance on the ba/pa contrast assessed via SDT.

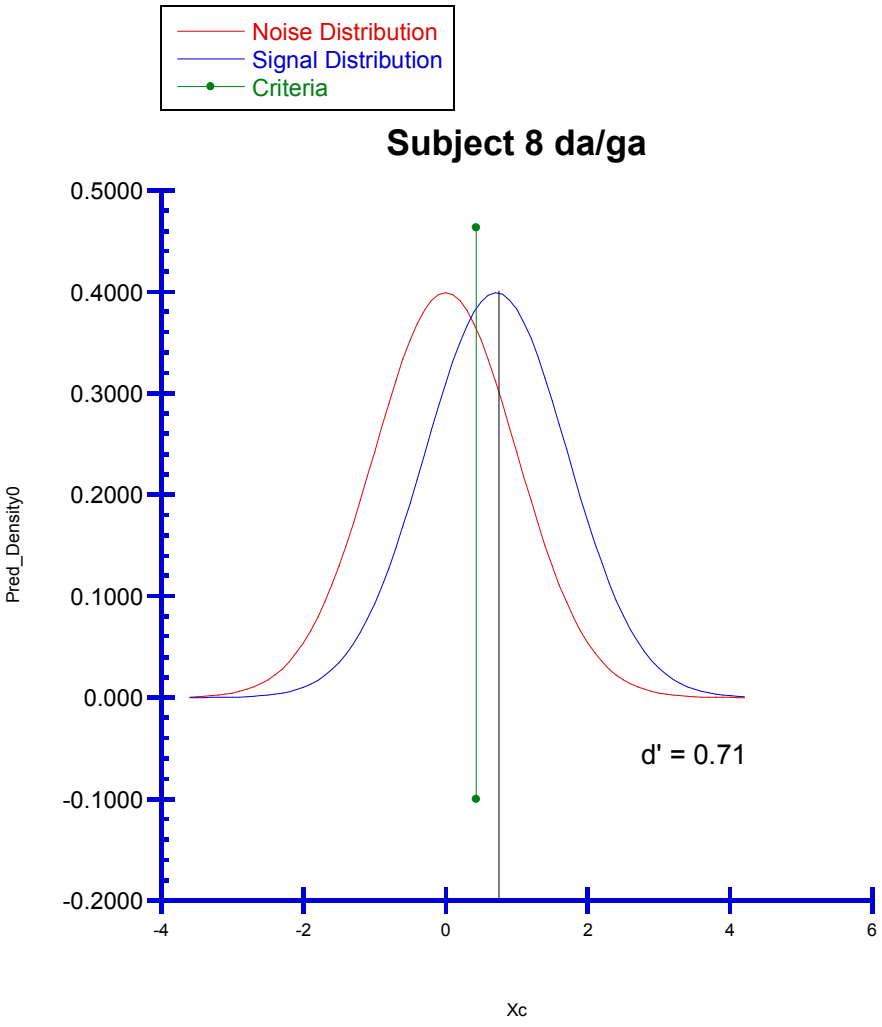


Figure 3. Subject 8's performance on the da/ga contrast assessed via SDT.

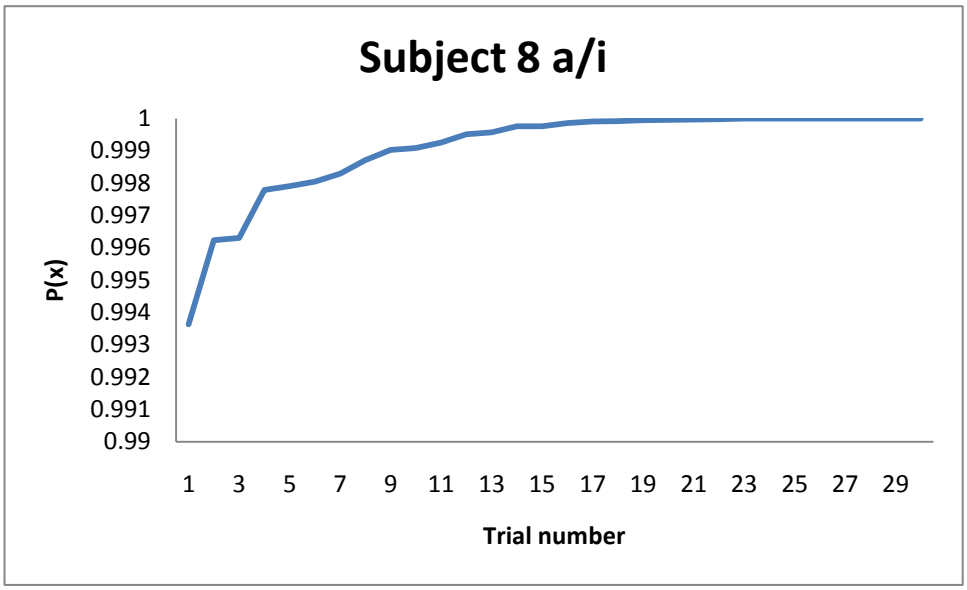


Figure 4. Cumulative distribution function for Subject 8's performance on the a/i contrast.

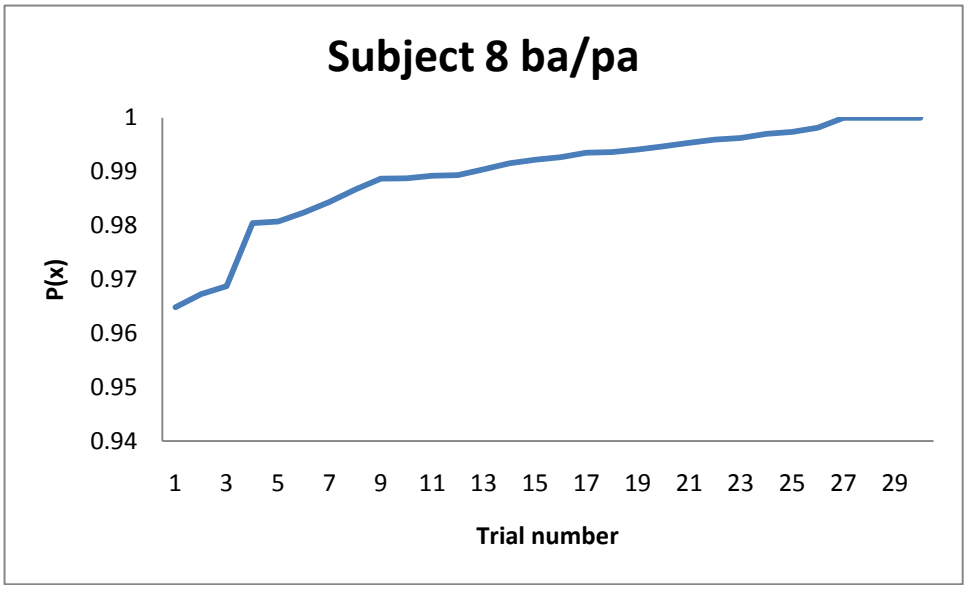


Figure 5. Cumulative distribution function for Subject 8's performance on the ba/pa contrast.

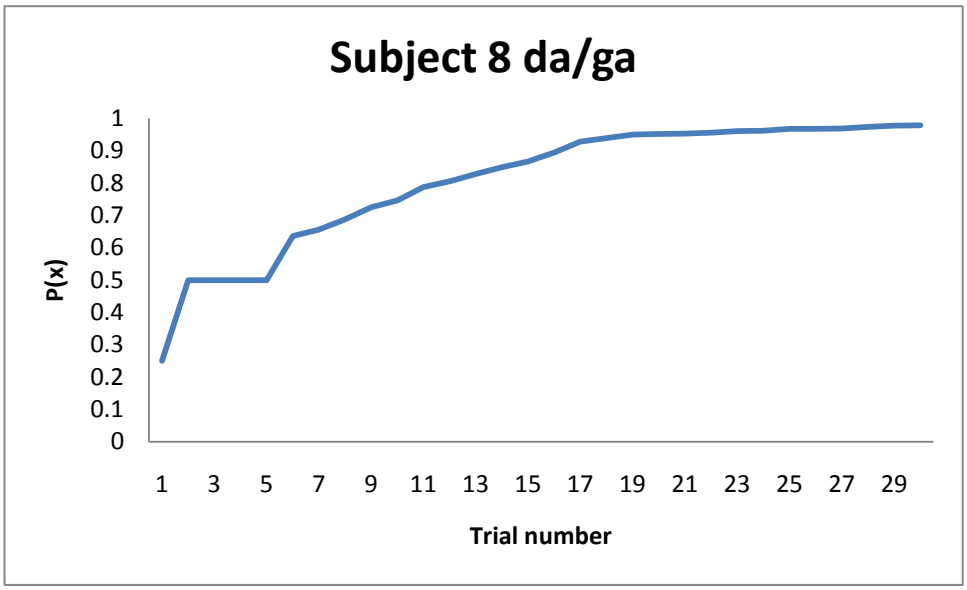


Figure 6. Cumulative distribution function for Subject 8's performance on the da/ga contrast.

Appendix I: Subject 9

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 80* | 0.79* | 0.67* | 1.64* | 11* | 9* |
| ba/pa | 76* | 0.78* | 0.61* | 1.56* | 11* | 9* |
| da/ga | 70* | 0.72* | 0.5* | 1.25* | 20* | 19* |

Table 1. Subject 9's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

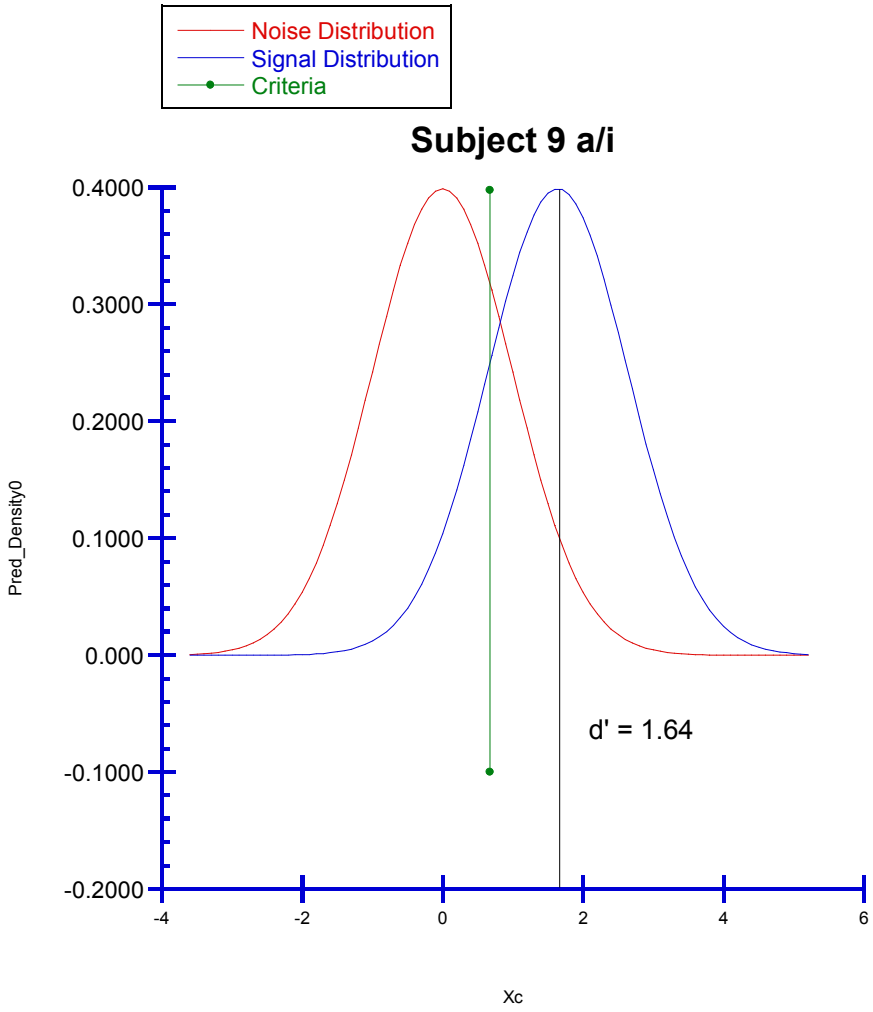


Figure 1. Subject 9's performance on the a/i contrast assessed via SDT.

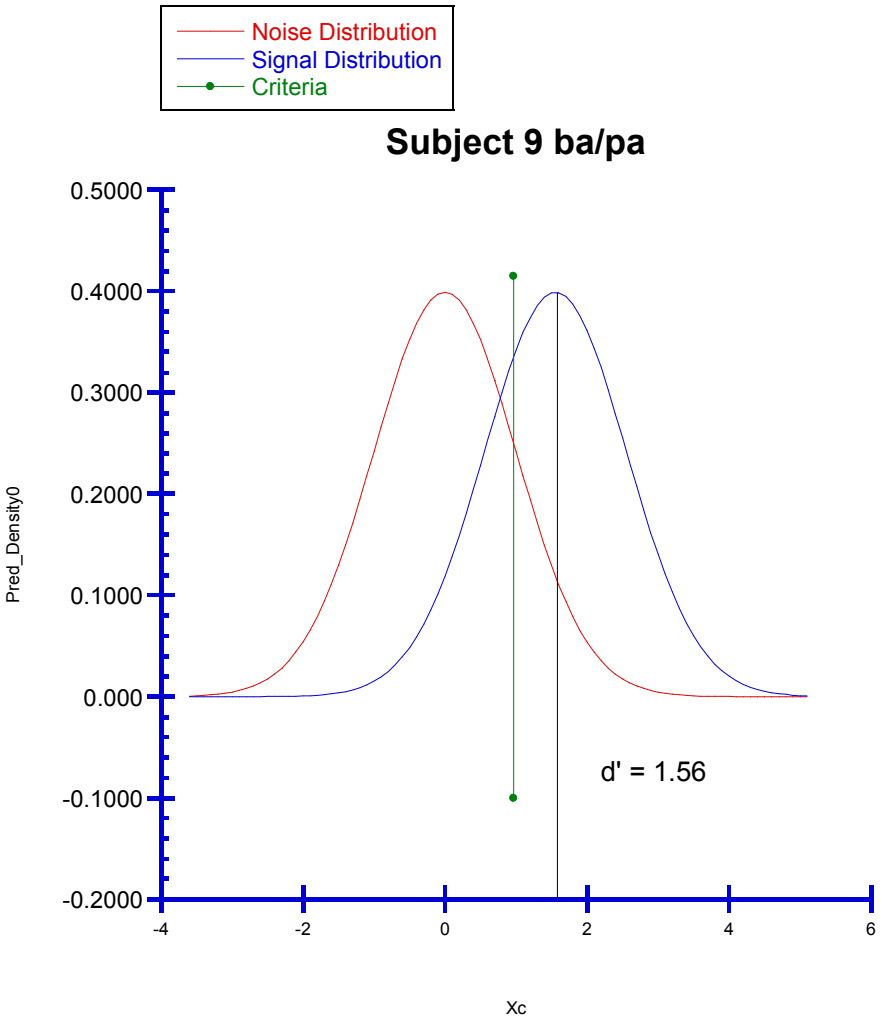


Figure 2. Subject 9's performance on the ba/pa contrast assessed via SDT.

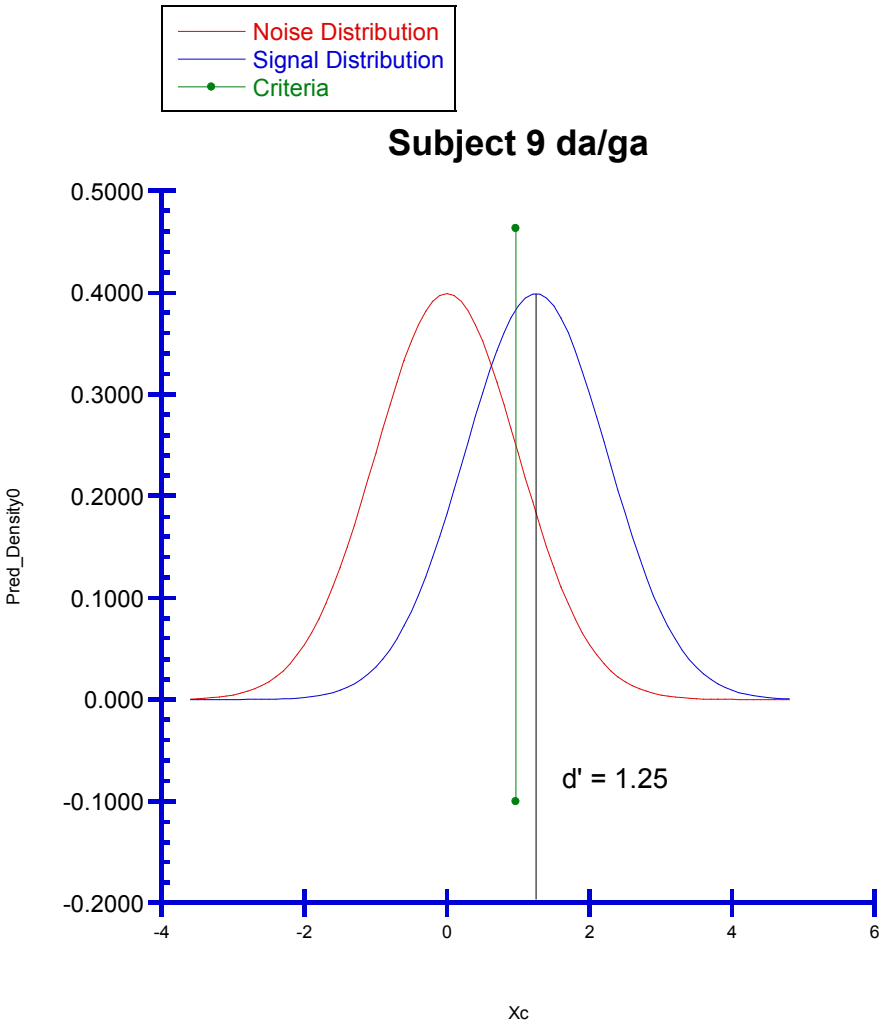


Figure 3. Subject 9's performance on the da/ga contrast assessed via SDT.

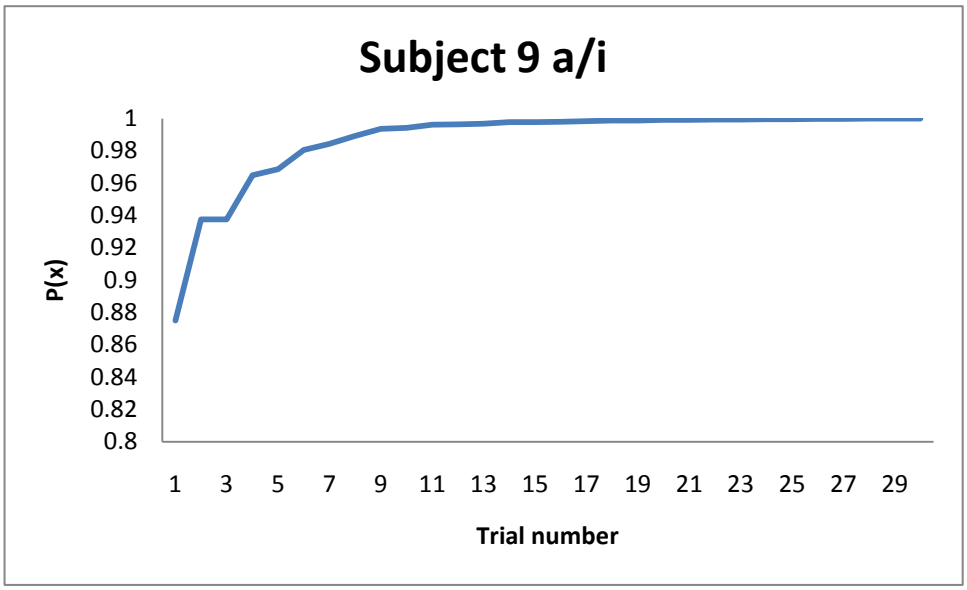


Figure 4. Cumulative distribution function for Subject 9's performance on the a/i contrast.

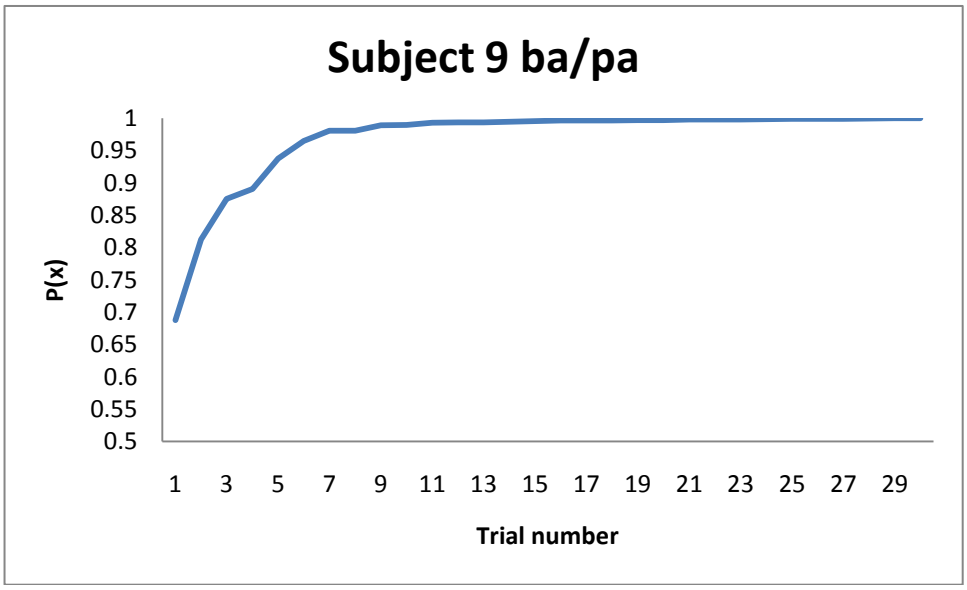


Figure 5. Cumulative distribution function for Subject 9's performance on the ba/pa contrast.

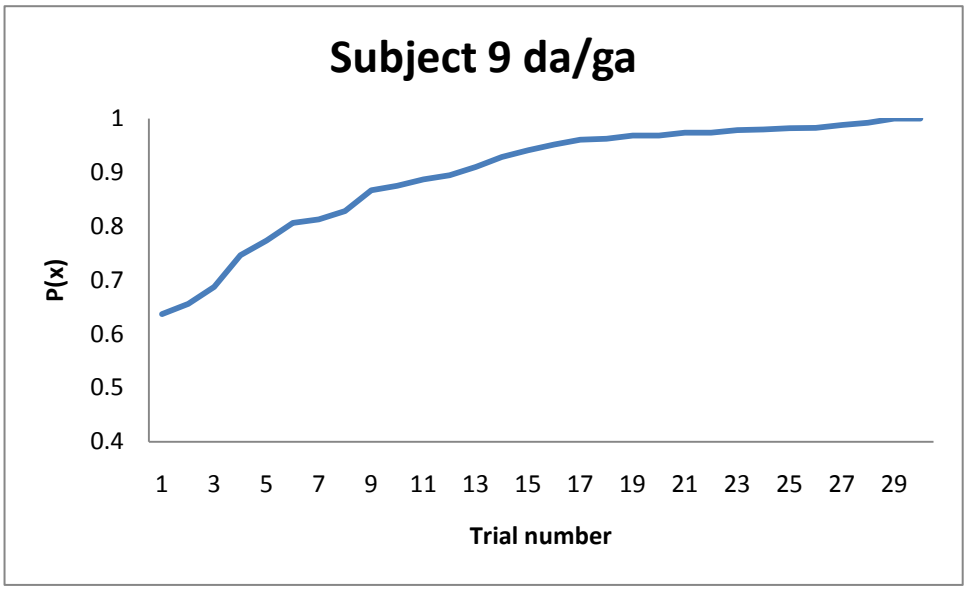


Figure 6. Cumulative distribution function for Subject 9's performance on the da/ga contrast.

Appendix J: Subject 10

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 53 | 0.58* | 0.22 | 0.54 | 30 | 30 |
| ba/pa | 57* | 0.63* | 0.28 | 0.95 | 21* | 30 |
| da/ga | 37 | 0.46 | -0.06 | -0.54 | 30 | 30 |

Table 1. Subject 10's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

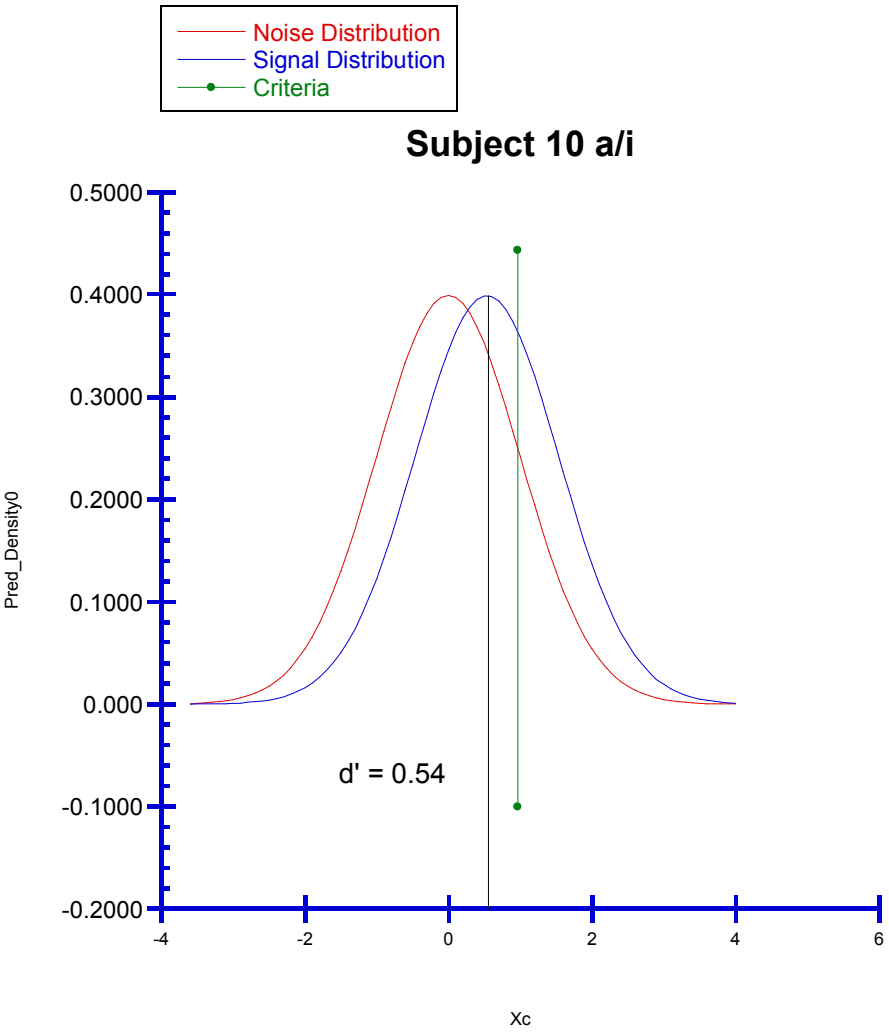


Figure 1. Subject 10's performance on the a/i contrast assessed via SDT.

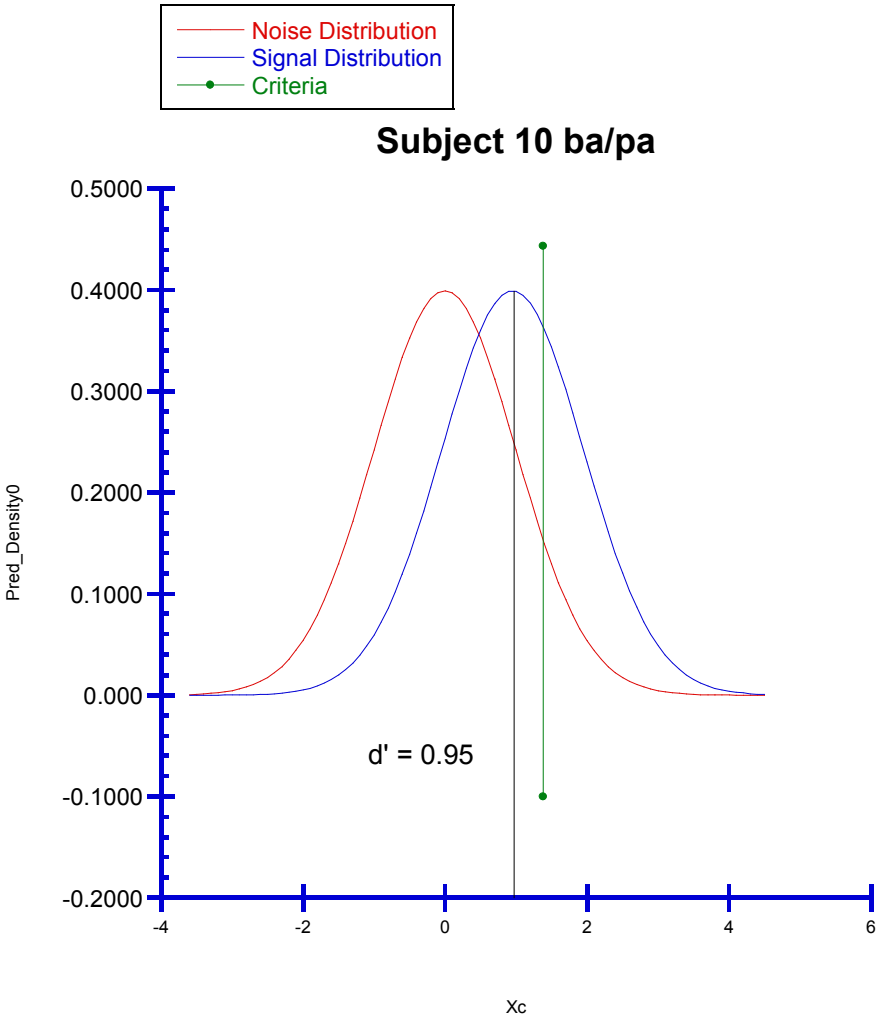


Figure 2. Subject 10's performance on the ba/pa contrast assessed via SDT.

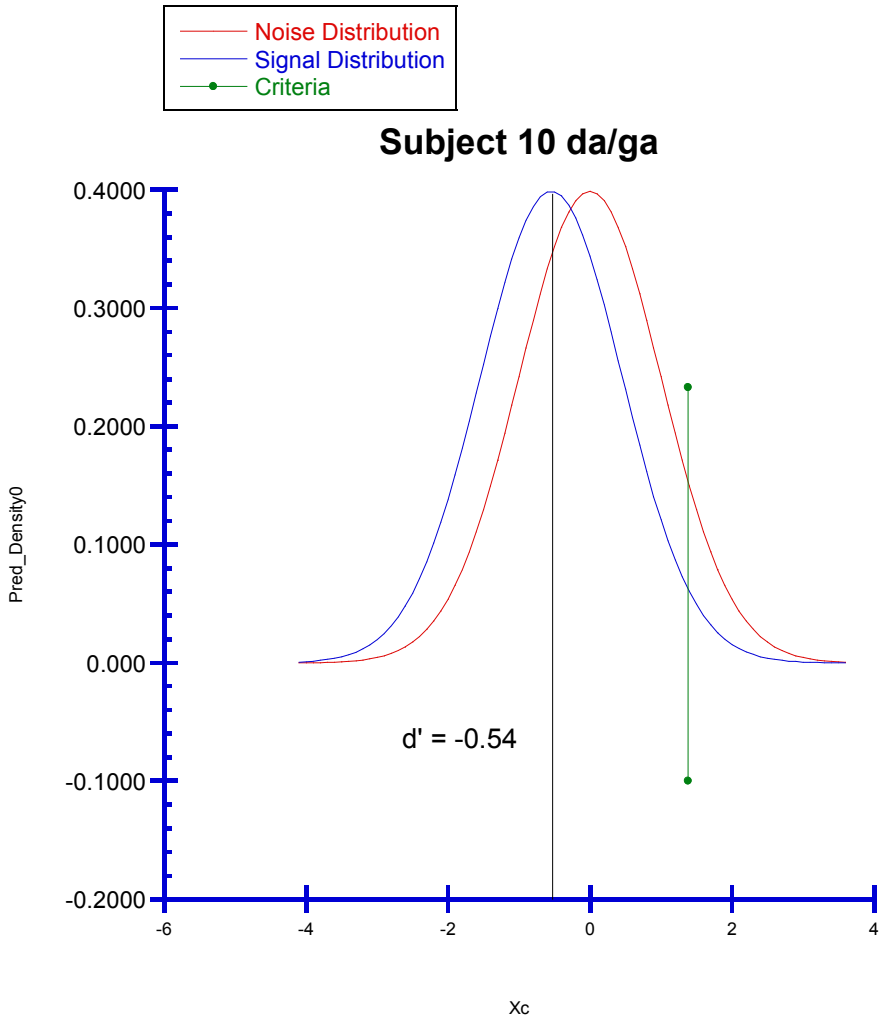


Figure 3. Subject 10's performance on the da/ga contrast assessed via SDT.

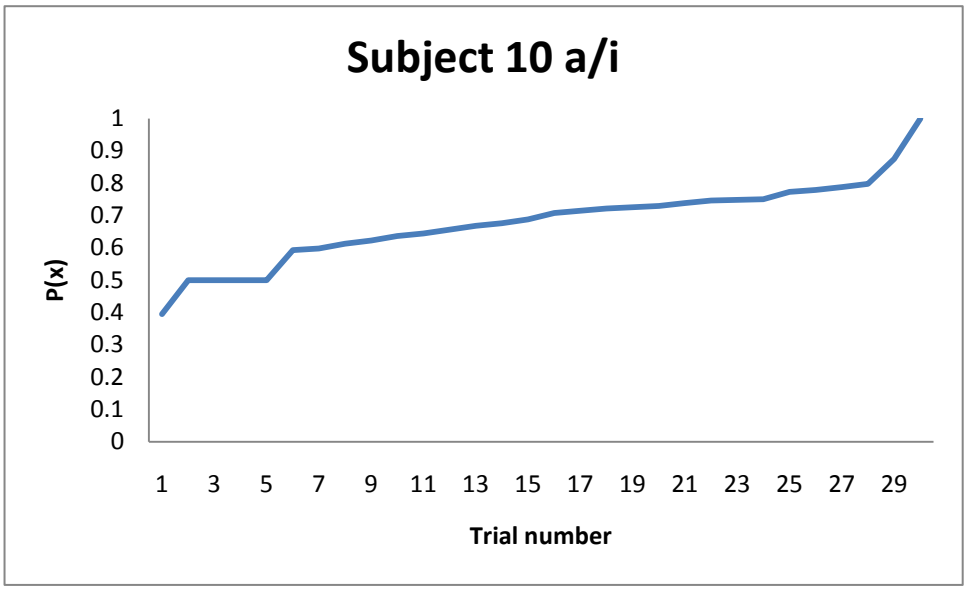


Figure 4. Cumulative distribution function for Subject 10's performance on the a/i contrast.

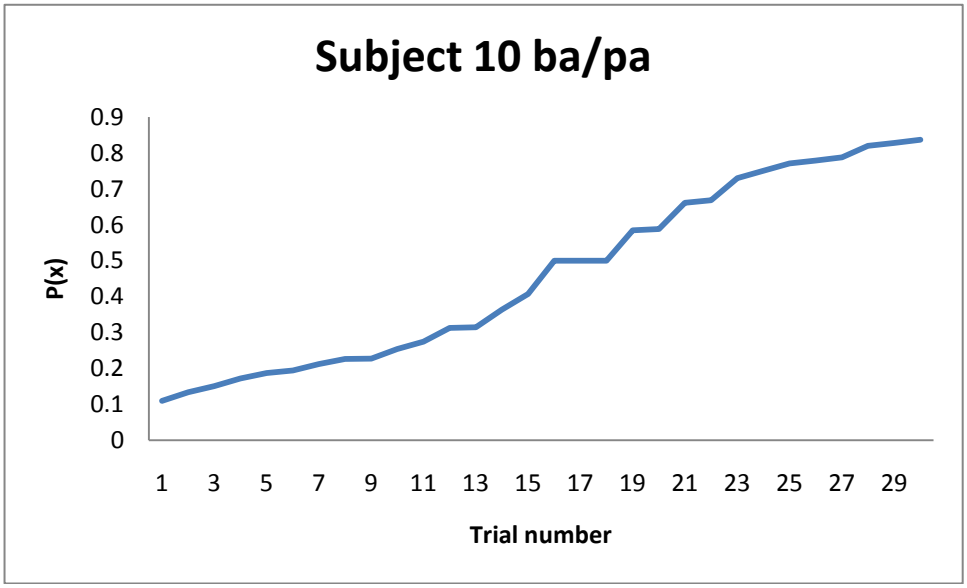


Figure 5. Cumulative distribution function for Subject 10's performance on the ba/pa contrast.

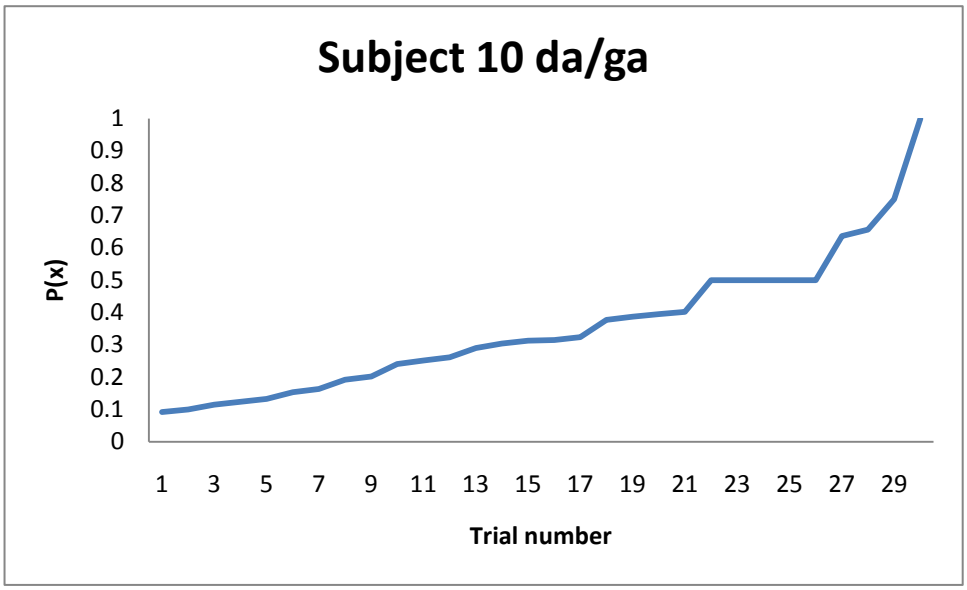


Figure 6. Cumulative distribution function for Subject 10's performance on the da/ga contrast.

Appendix K: Subject 11

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 77* | 0.75* | 0.61* | 1.4* | 22* | 12* |
| ba/pa | 87* | 0.83* | 0.78* | 2.36* | 14* | 9* |
| da/ga | 47 | 0.53 | 0.11 | 0.2 | 30 | 30 |

Table 1. Subject 11's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

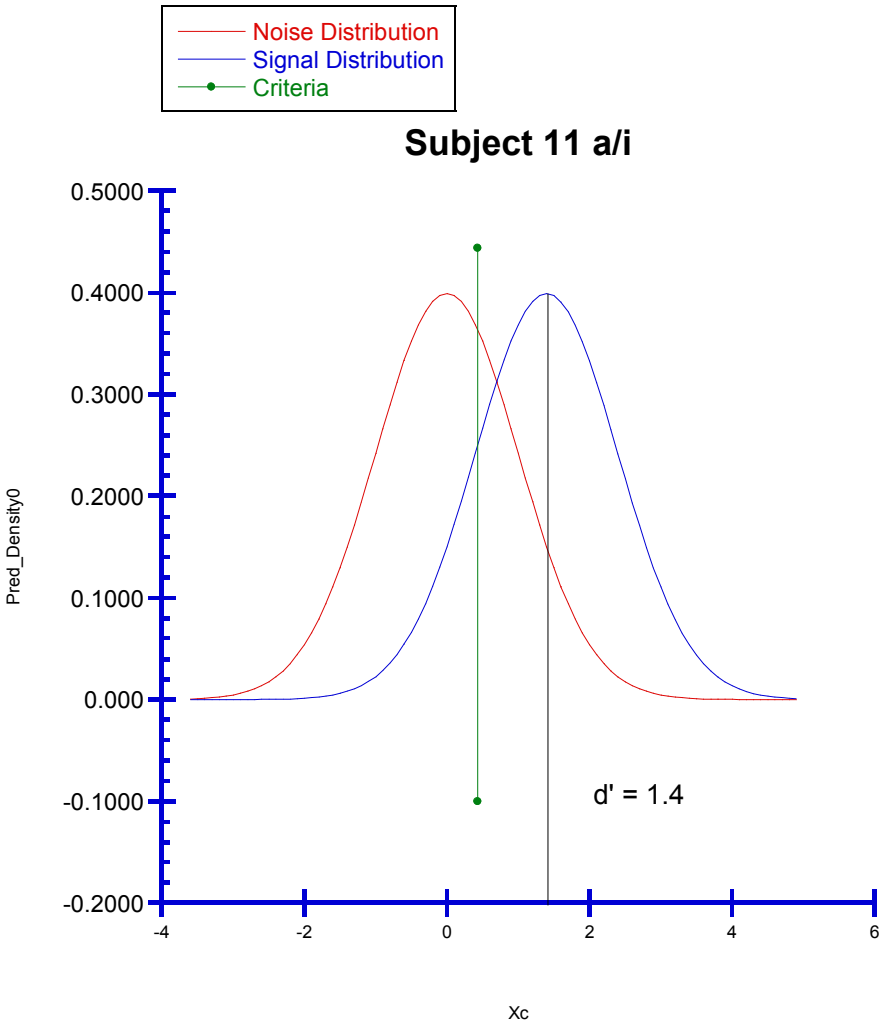


Figure 1. Subject 11's performance on the a/i contrast assessed via SDT.

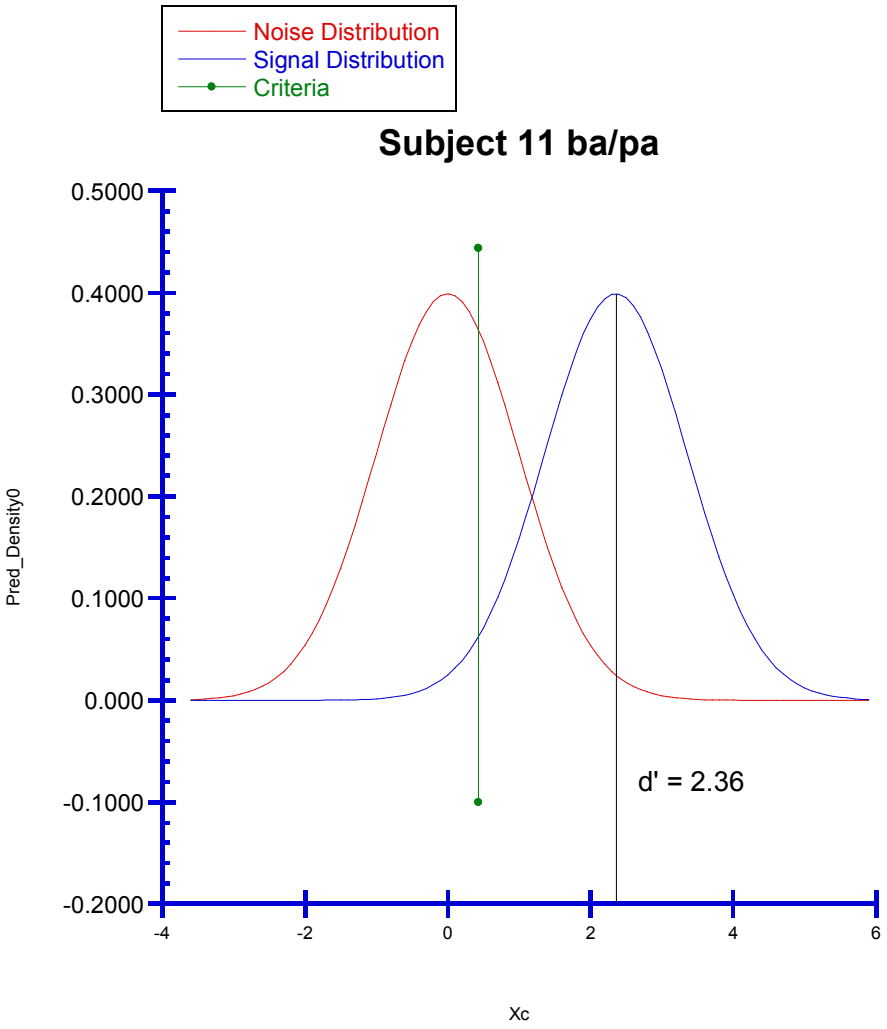


Figure 2. Subject 11's performance on the ba/pa contrast assessed via SDT.

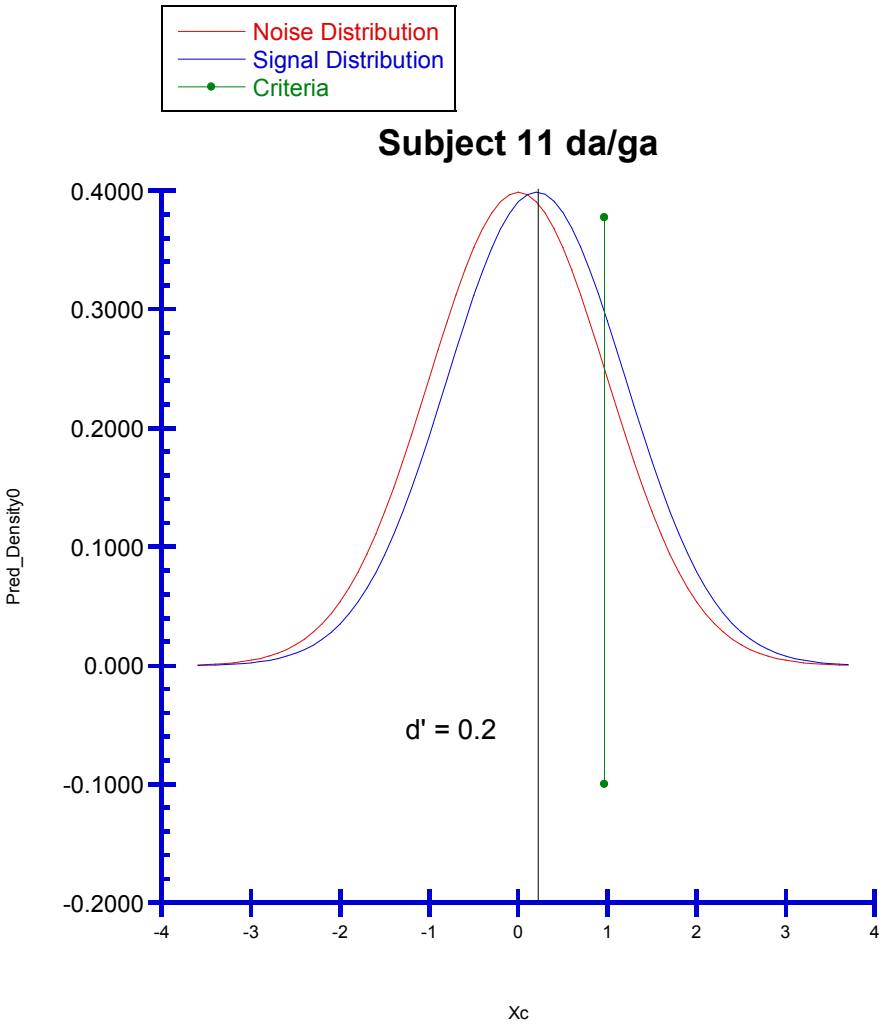


Figure 3. Subject 11's performance on the da/ga contrast assessed via SDT.

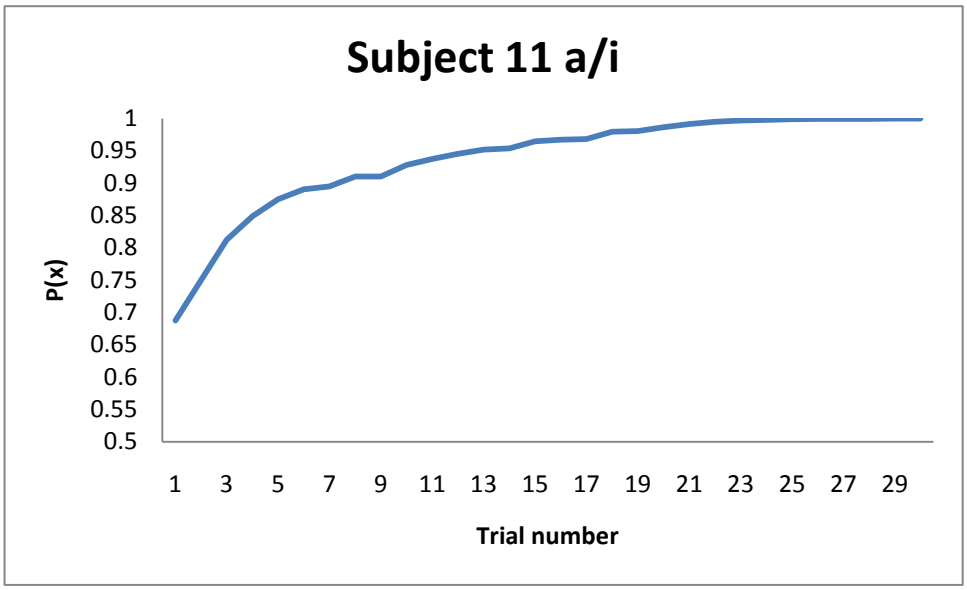


Figure 4. Cumulative distribution function for Subject 11's performance on the a/i contrast.

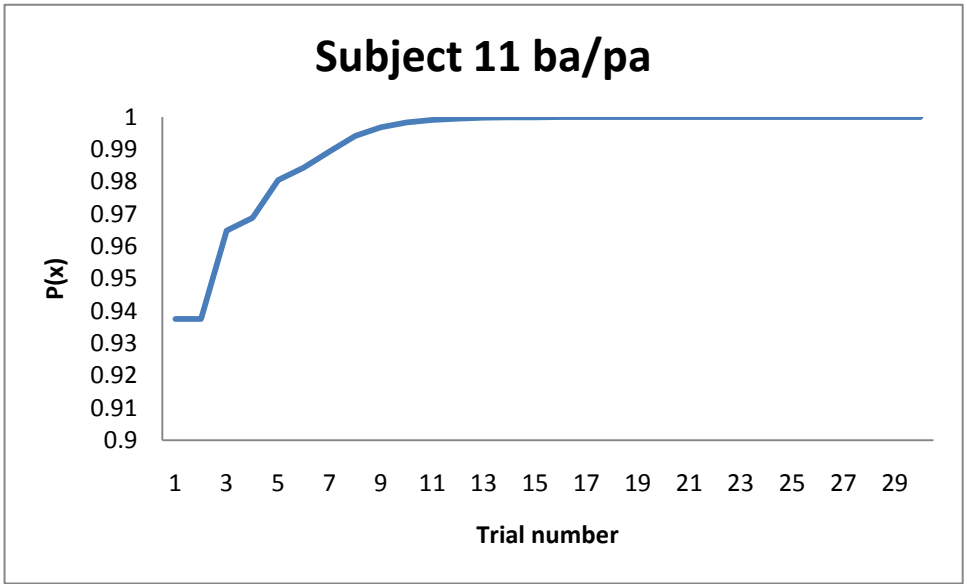


Figure 5. Cumulative distribution function for Subject 11's performance on the ba/pa contrast.

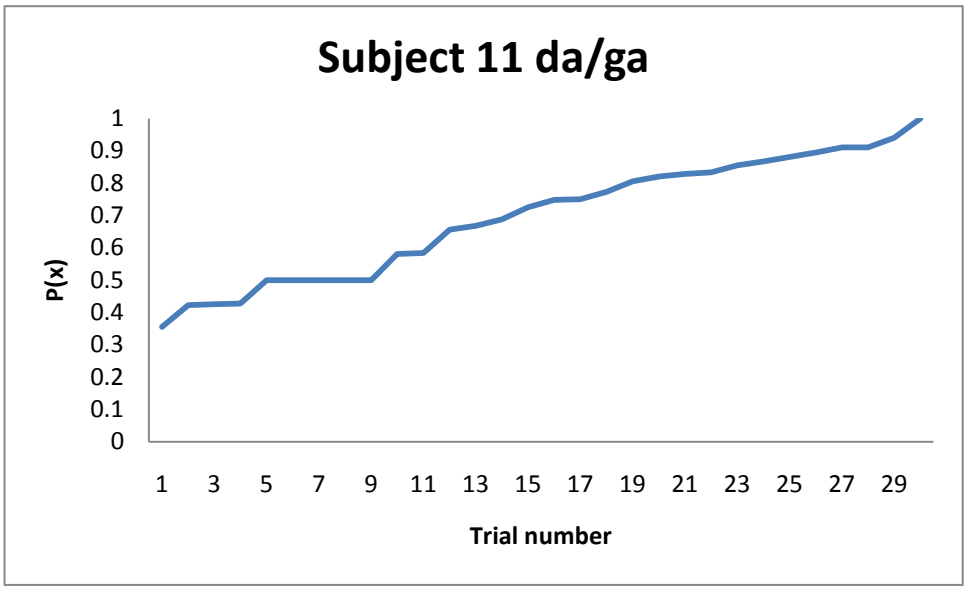


Figure 6. Cumulative distribution function for Subject 11's performance on the da/ga contrast.

Appendix L: Subject 12

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 83* | 0.82* | 0.72* | 1.9* | 8* | 8* |
| ba/pa | 70* | 0.7* | 0.5 | 1.02* | 17* | 12* |
| da/ga | 70* | 0.72* | 0.5* | 1.25* | 22* | 22* |

Table 1. Subject 12's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

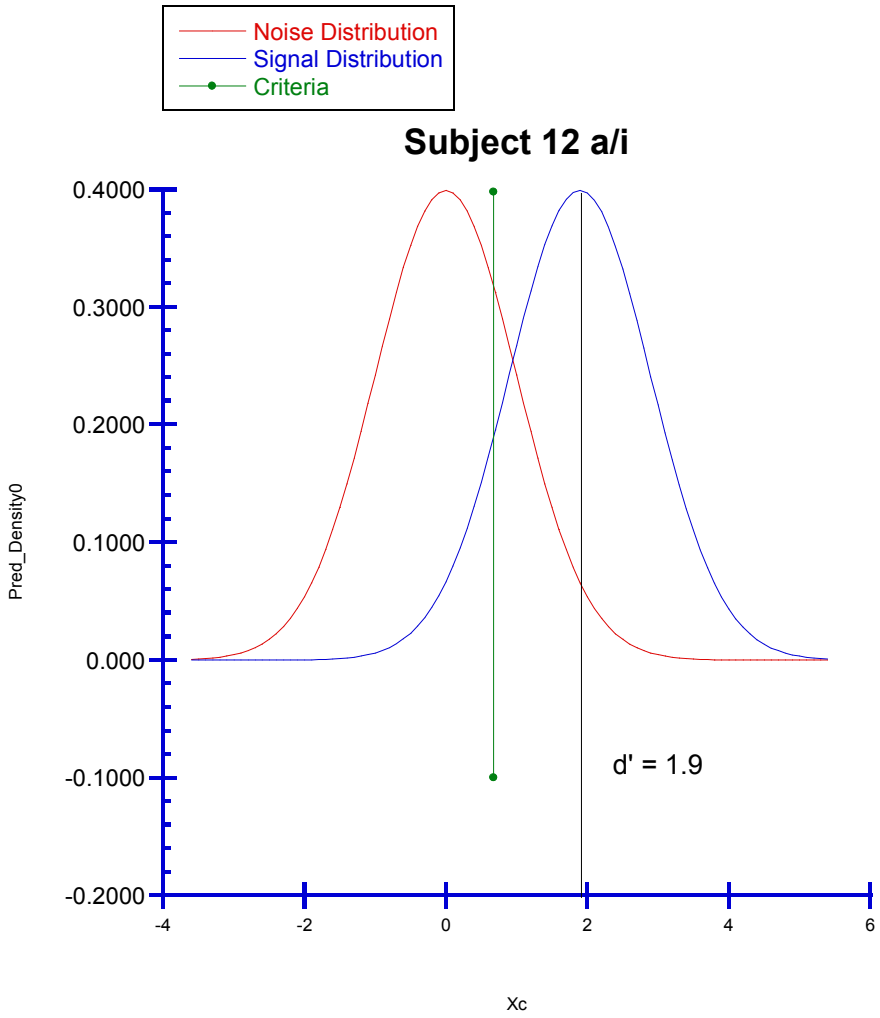


Figure 1. Subject 12's performance on the a/i contrast assessed via SDT.

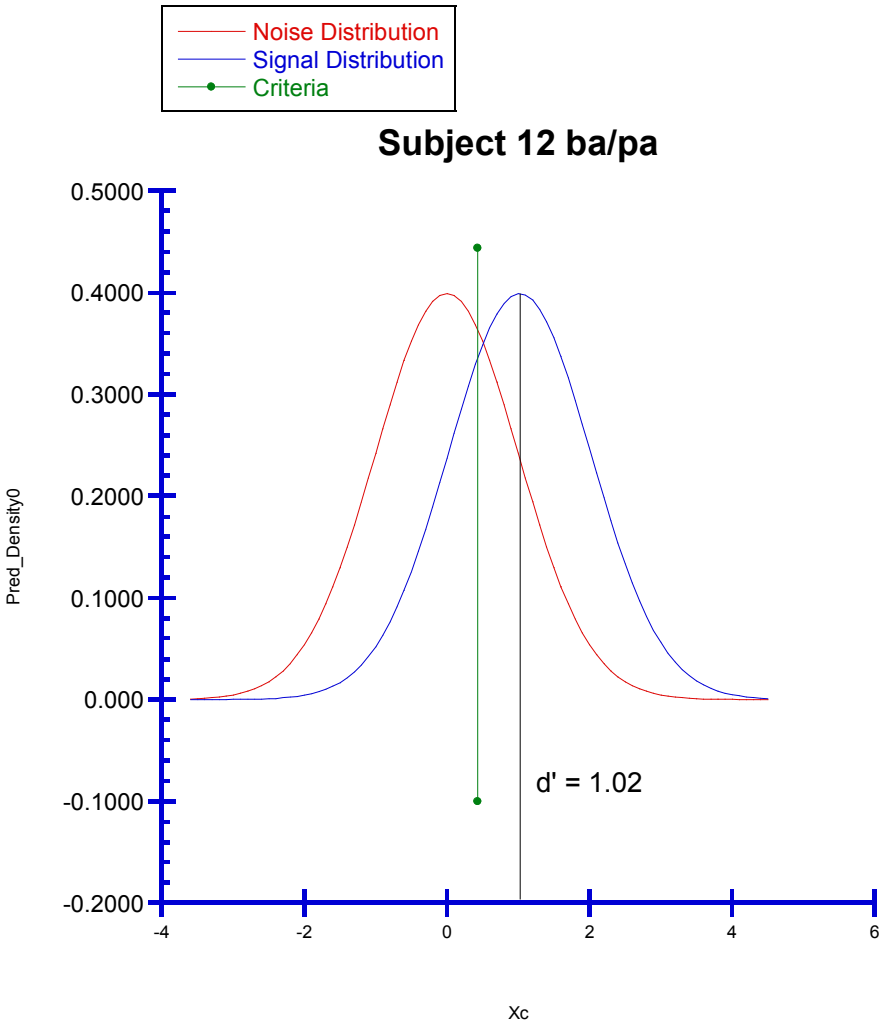


Figure 2. Subject 12's performance on the ba/pa contrast assessed via SDT.

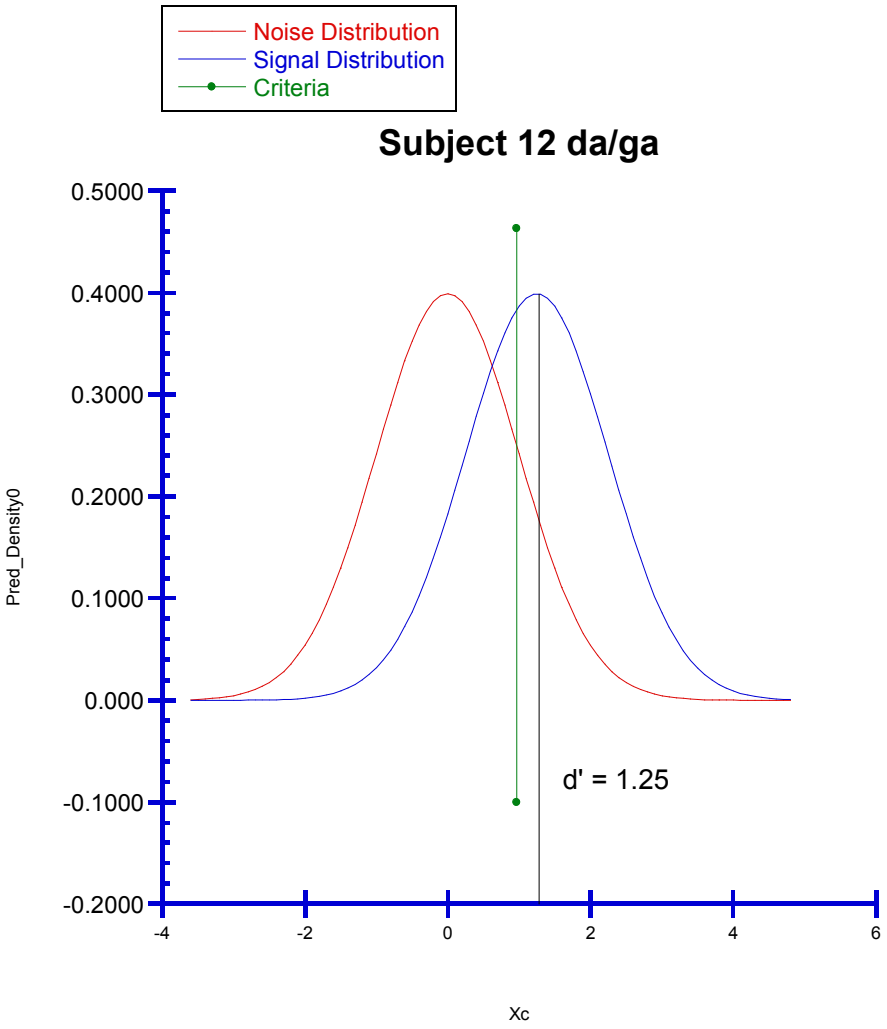


Figure 3. Subject 12's performance on the da/ga contrast assessed via SDT.

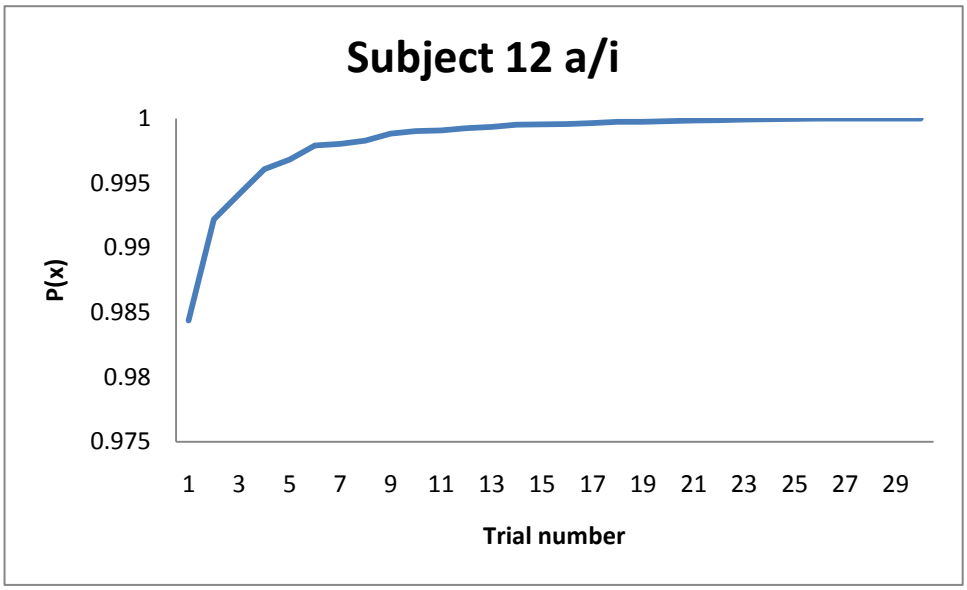


Figure 4. Cumulative distribution function for Subject 12's performance on the a/i contrast.

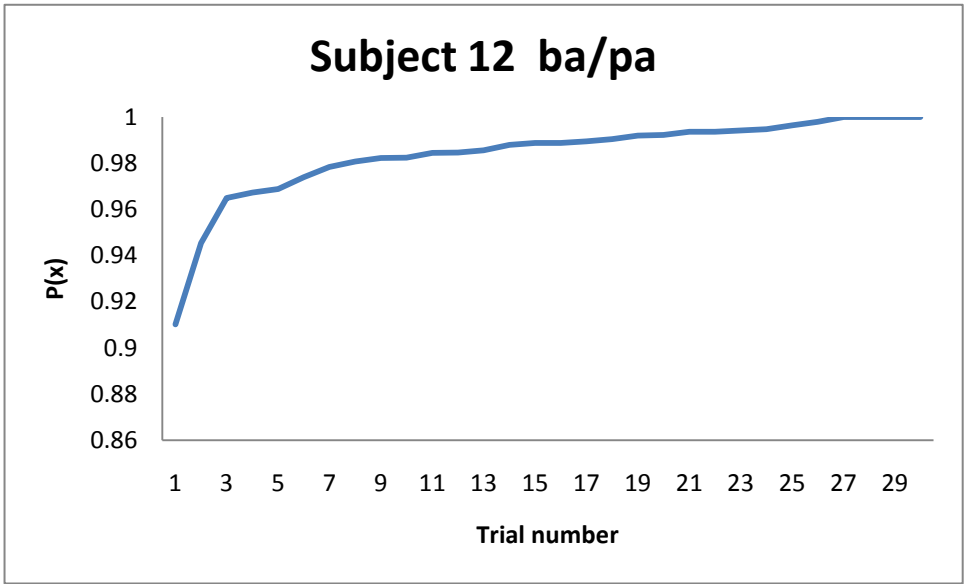


Figure 5. Cumulative distribution function for Subject 12's performance on the ba/pa contrast.

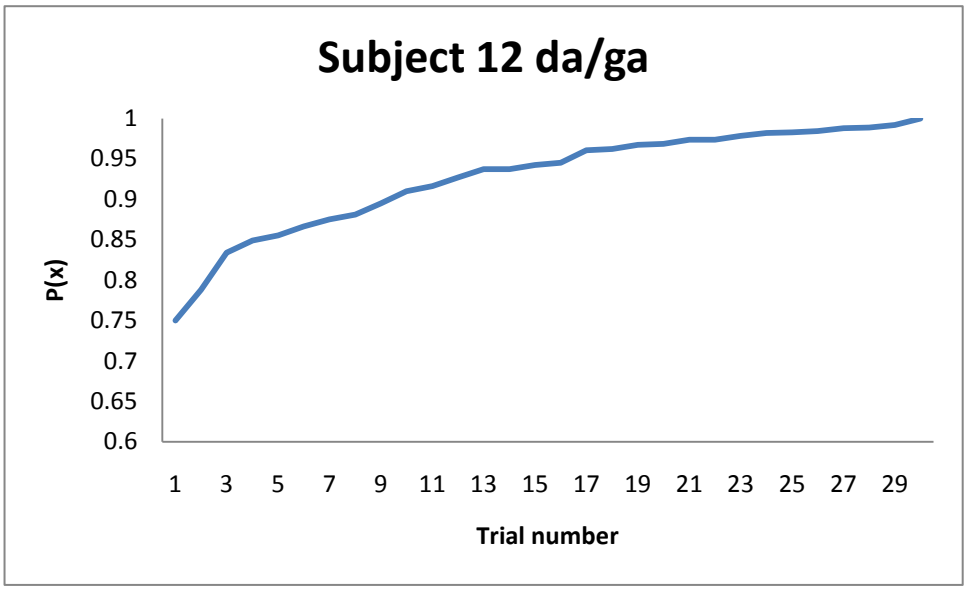


Figure 6. Cumulative distribution function for Subject 12's performance on the da/ga contrast.

Appendix M: Subject 13

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 80* | 0.81* | 0.67* | 1.73* | 11* | 9* |
| ba/pa | 70* | 0.72* | 0.5 | 1.25* | 15* | 9* |
| da/ga | 70* | 0.75* | 0.5* | 1.75* | 14* | 17* |

Table 1. Subject 13's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

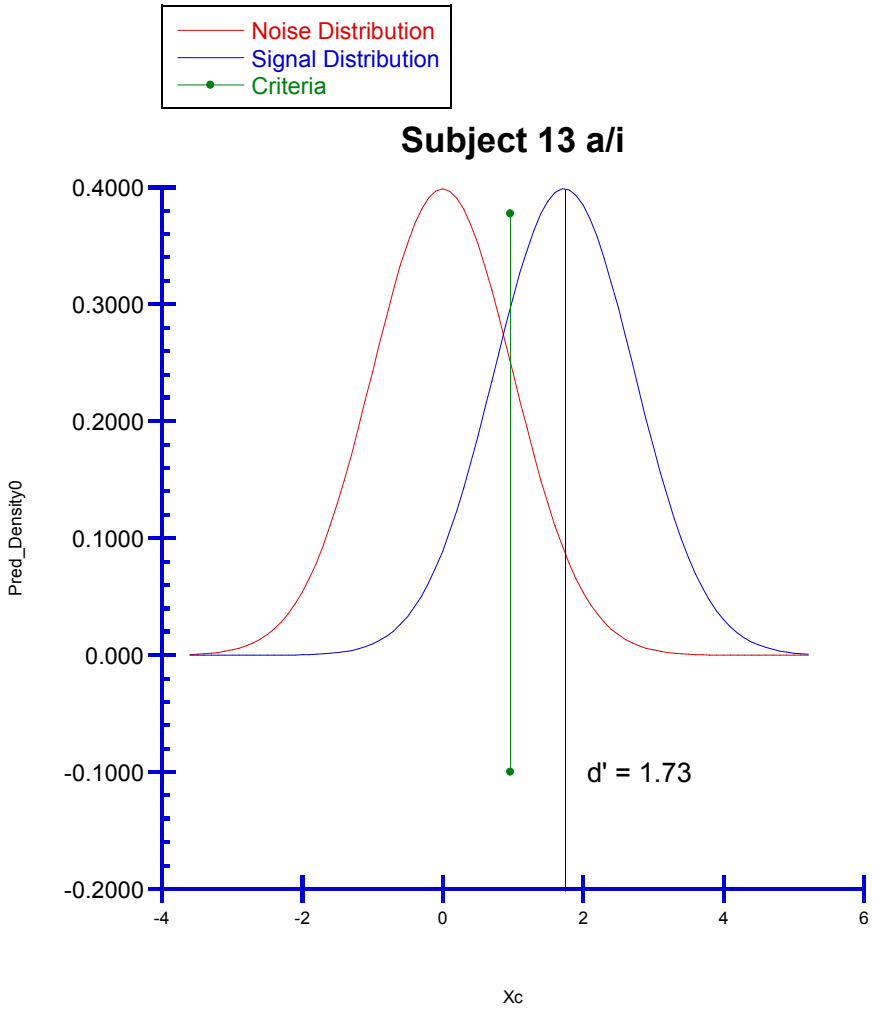


Figure 1. Subject 13's performance on the a/i contrast assessed via SDT.

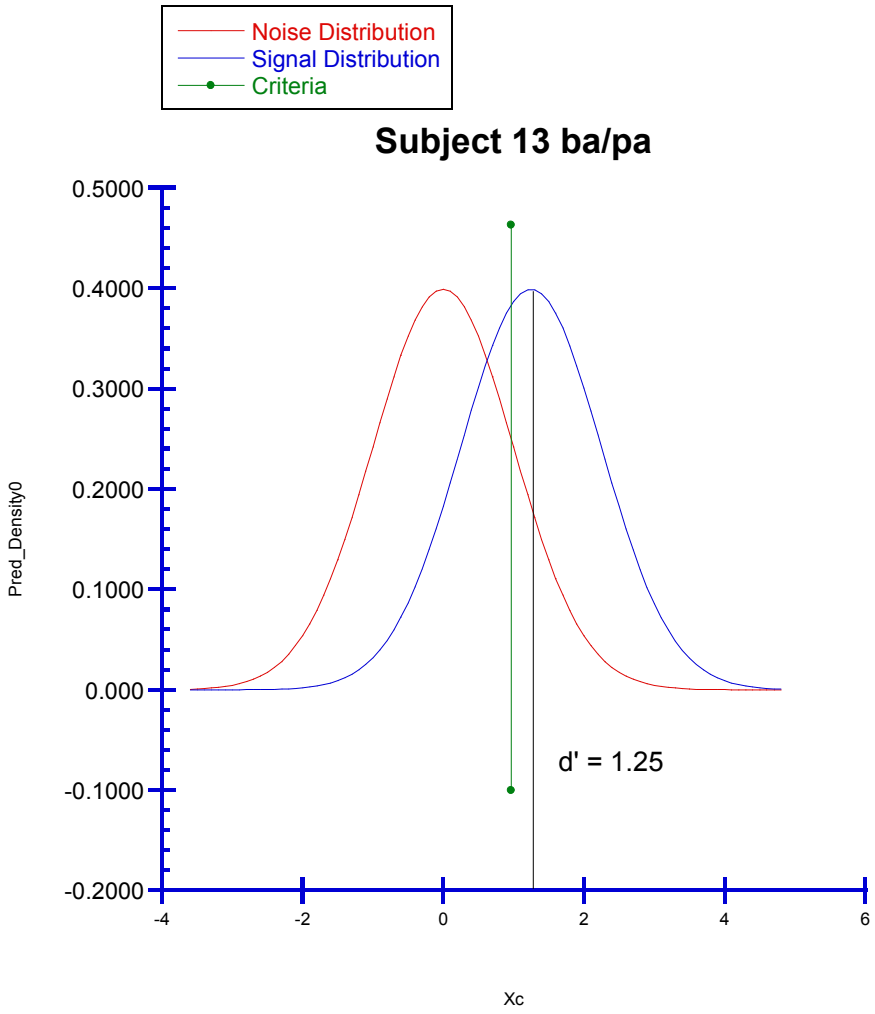


Figure 2. Subject 13's performance on the ba/pa contrast assessed via SDT.

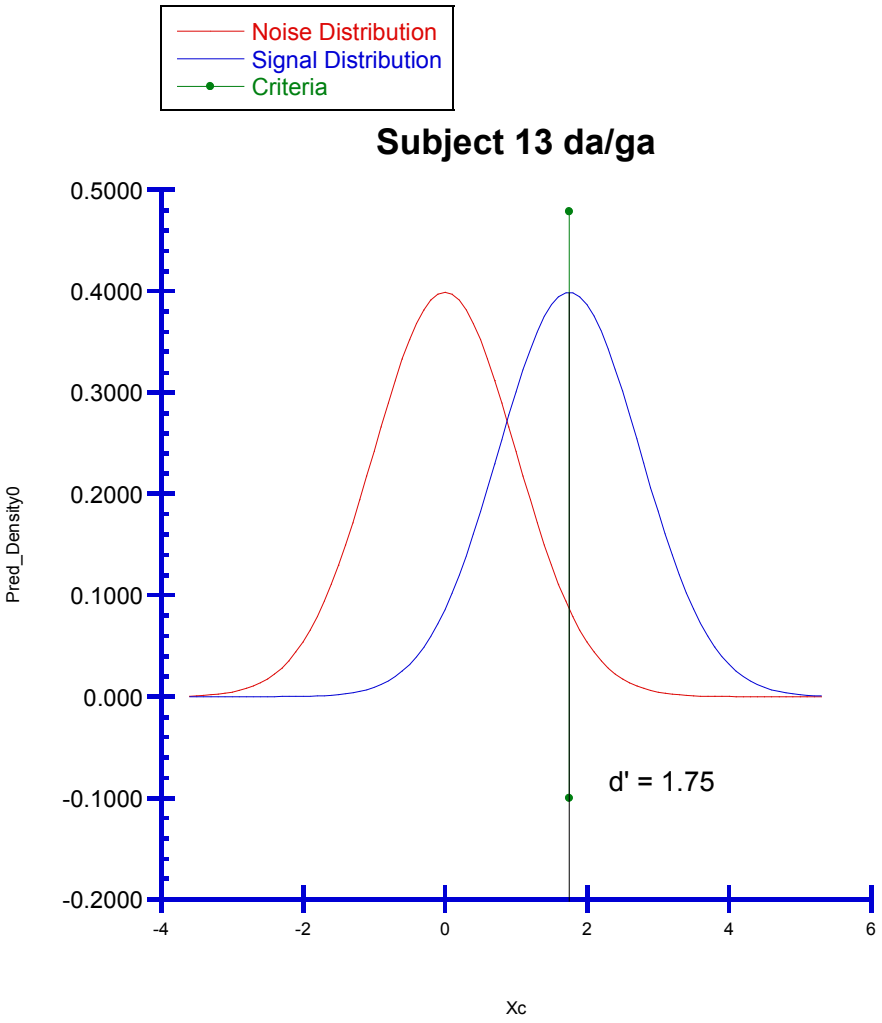


Figure 3. Subject 13's performance on the da/ga contrast assessed via SDT.

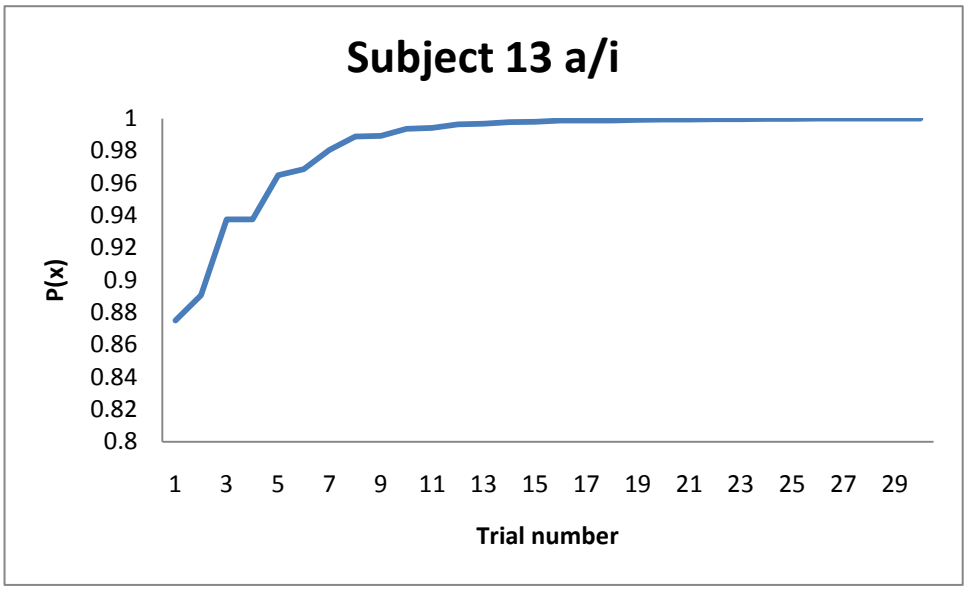


Figure 4. Cumulative distribution function for Subject 13's performance on the a/i contrast.

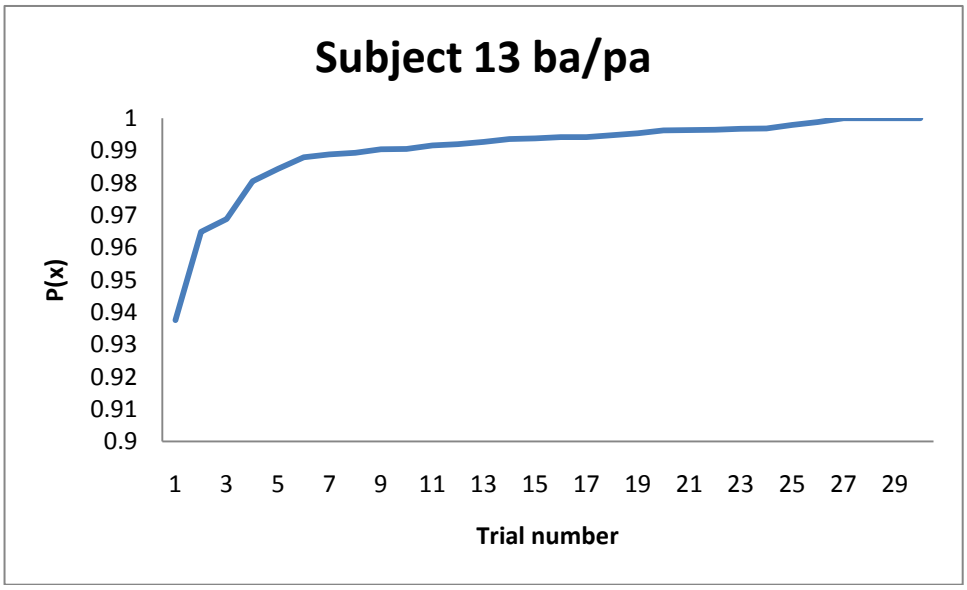


Figure 5. Cumulative distribution function for Subject 13's performance on the ba/pa contrast.

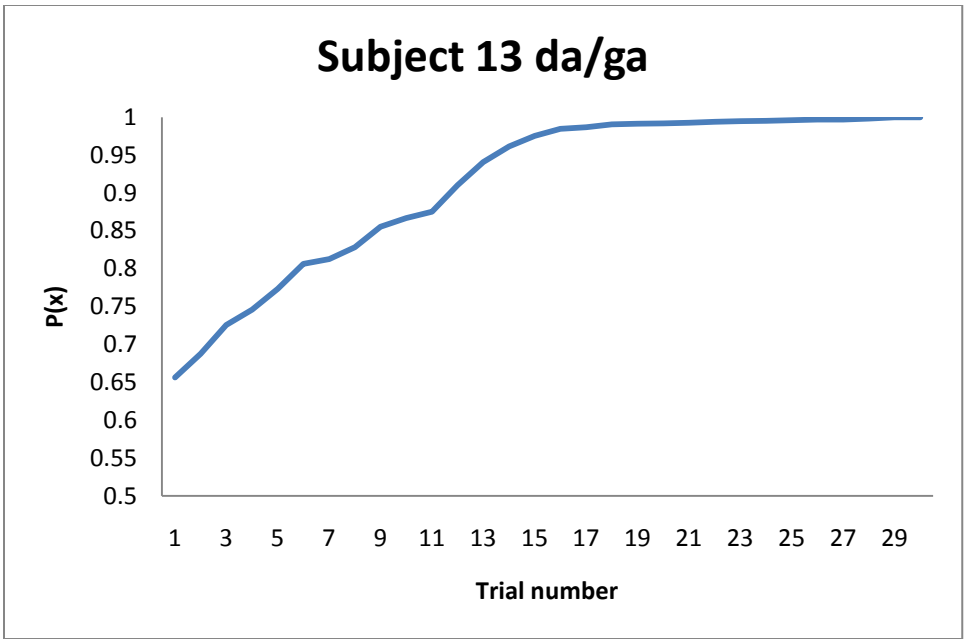


Figure 6. Cumulative distribution function for Subject 13's performance on the da/ga contrast.

Appendix N: Subject 14

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 90* | 0.89* | 0.83* | 2.56* | 8* | 8* |
| ba/pa | 77* | 0.78* | 0.61* | 1.56* | 8* | 8* |
| da/ga | 63* | 0.66* | 0.39 | 0.81* | 30 | 19* |

Table 1. Subject 14's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

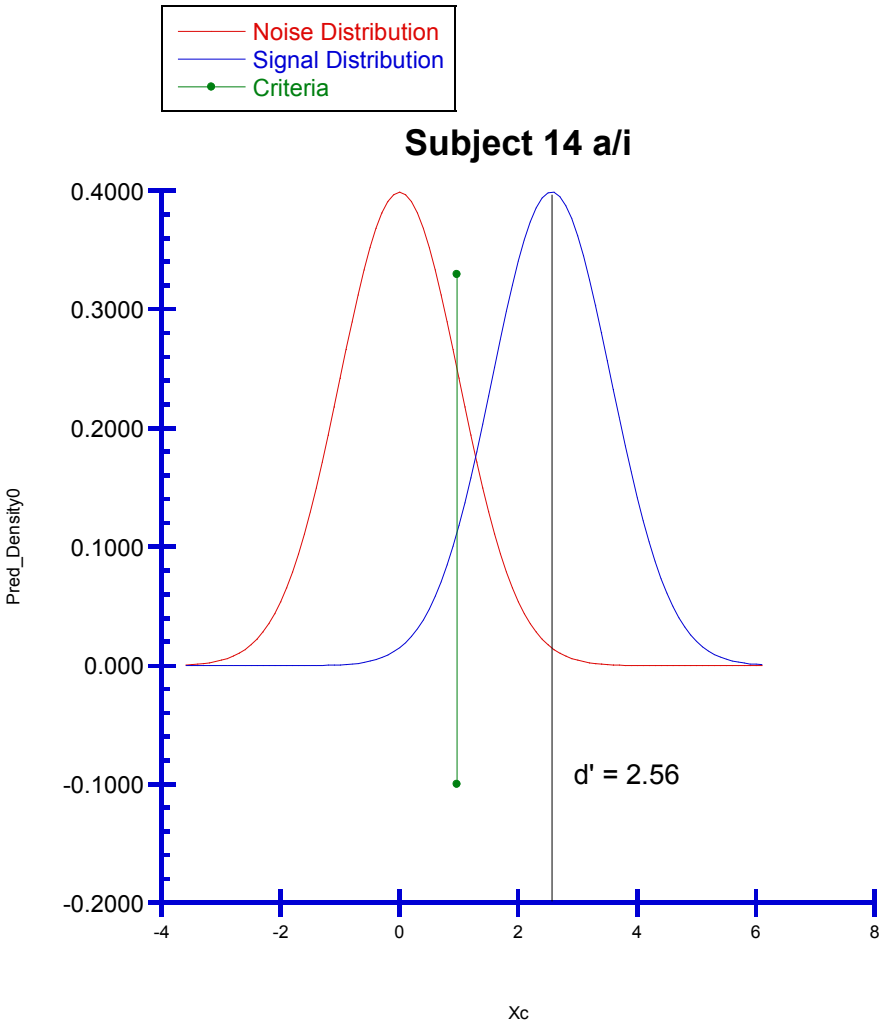


Figure 1. Subject 14's performance on the a/i contrast assessed via SDT.

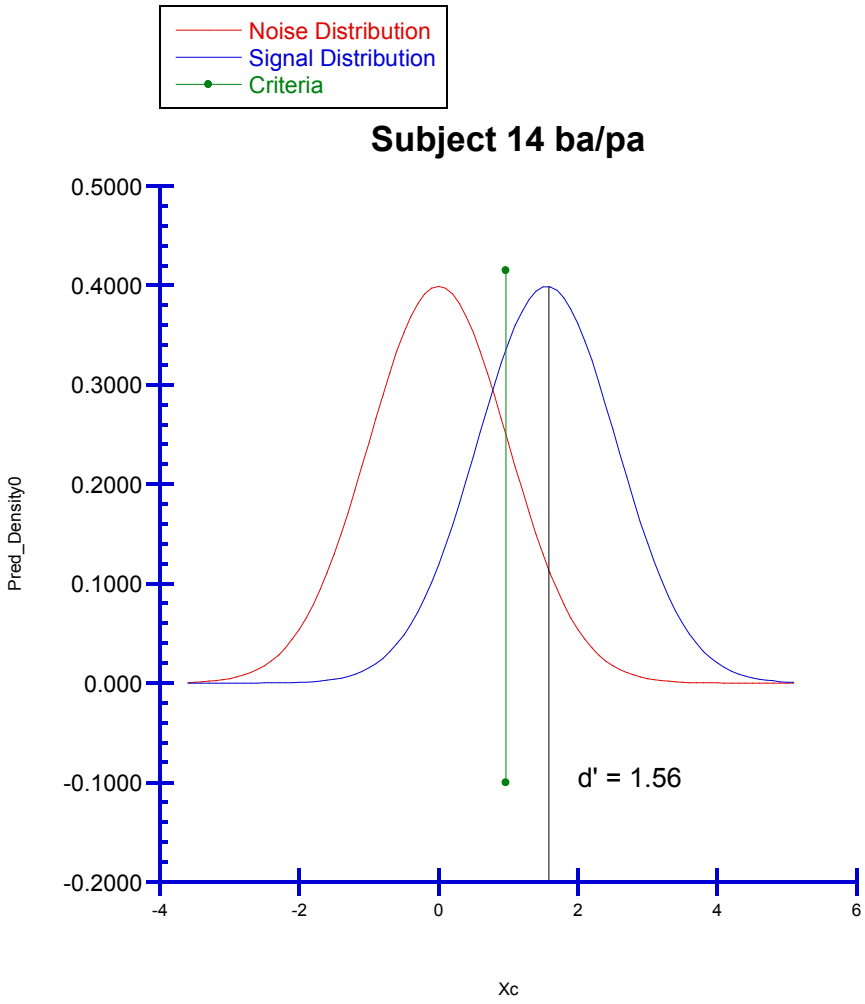


Figure 2. Subject 14's performance on the ba/pa contrast assessed via SDT.

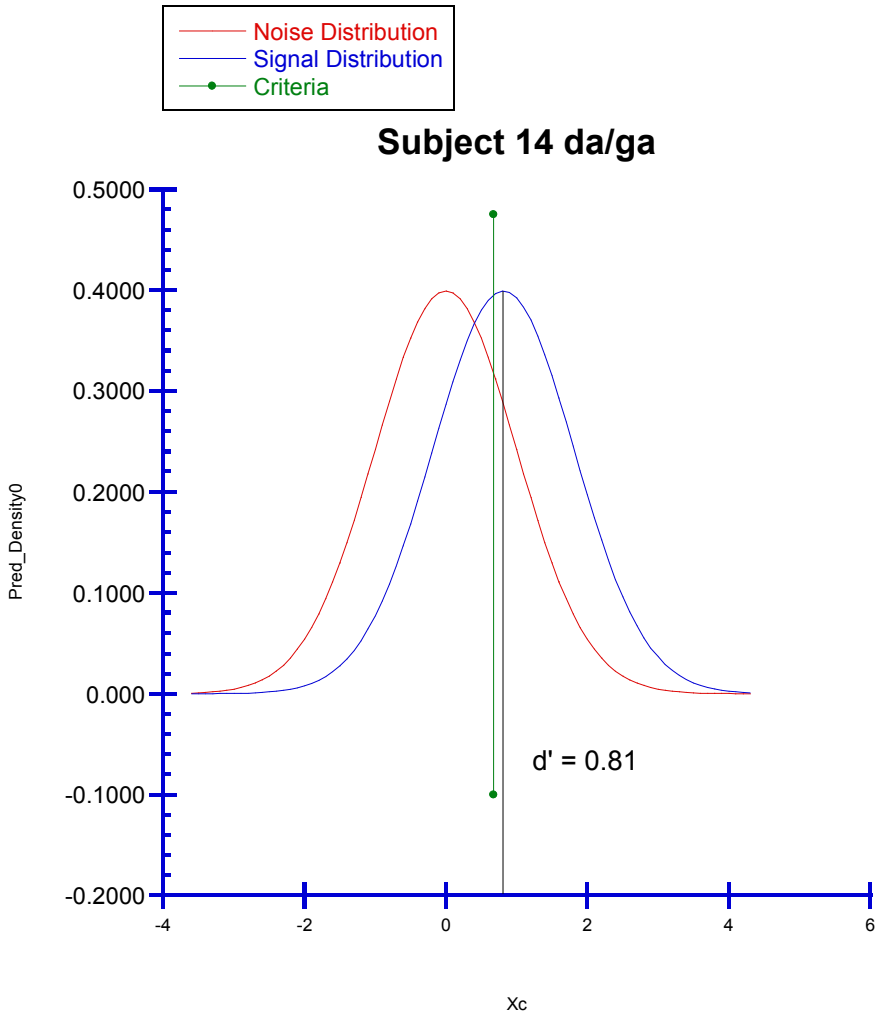


Figure 3. Subject 14's performance on the da/ga contrast assessed via SDT.

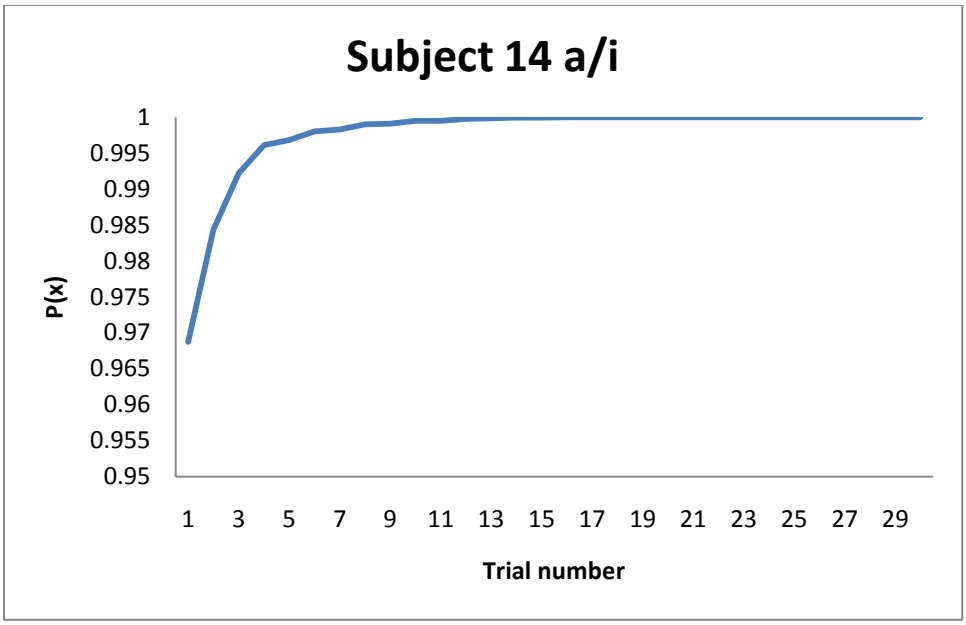


Figure 4. Cumulative distribution function for Subject 14's performance on the a/i contrast.

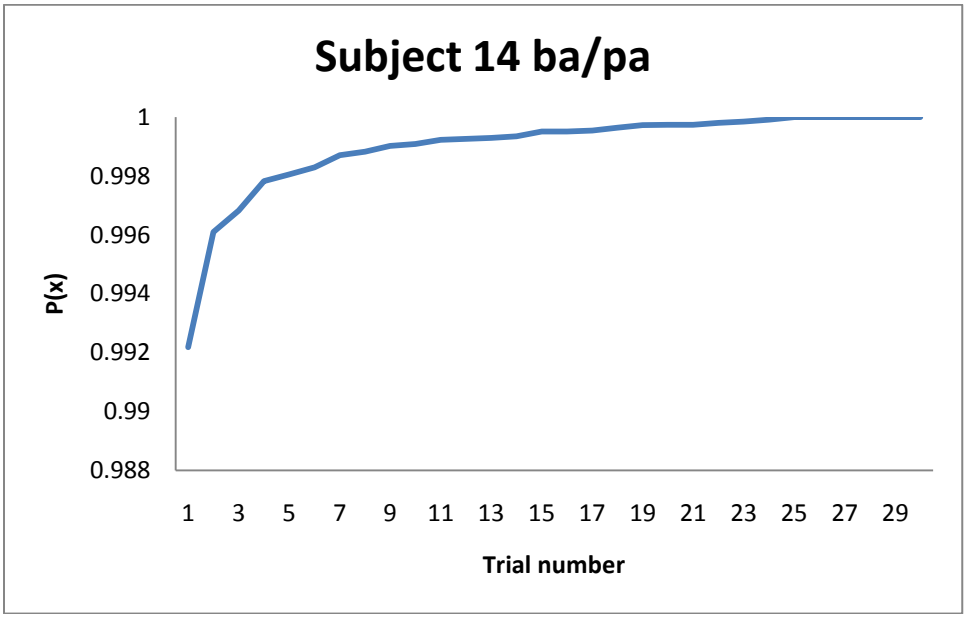


Figure 5. Cumulative distribution function for Subject 14's performance on the ba/pa contrast.

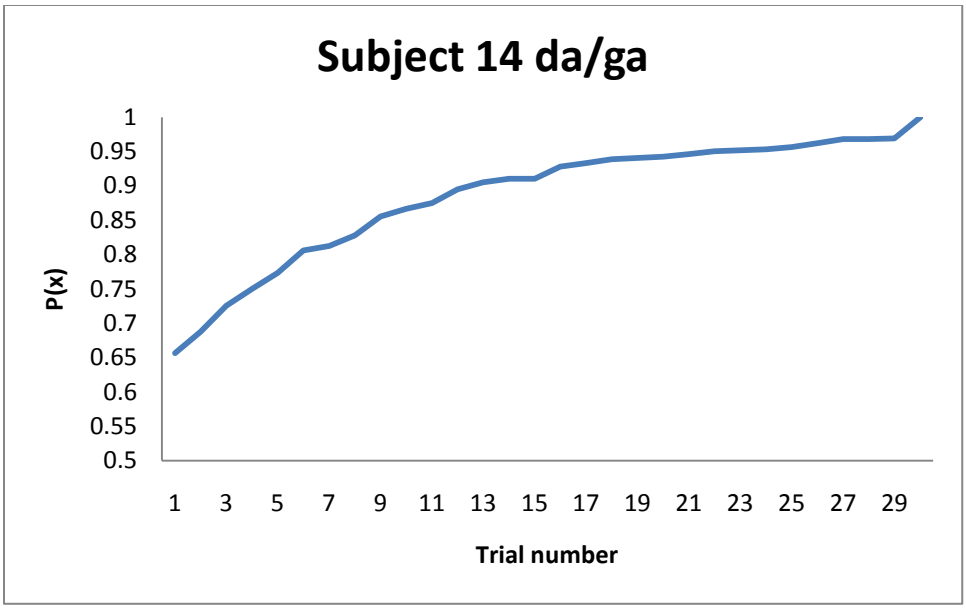


Figure 6. Cumulative distribution function for Subject 14's performance on the da/ga contrast.

Appendix O: Subject 15

| | percent correct | proportion correct | discrimination index | d' | criterion (7/8) | criterion (binomial probability) |
|---|---|---|---|---|---|---|
| a/i | 83* | 0.82* | 0.72* | 1.9* | 8* | 8* |
| ba/pa | 77* | 0.78* | 0.61* | 1.56* | 13* | 9* |
| da/ga | 70* | 0.72* | 0.5* | 1.25* | 9* | 8* |

Table 1. Subject 15's performance on each of the contrasts assessed via the different performance analysis methods. An asterisk (*) beside an outcome indicates passing performance based on that method's analysis requirements.

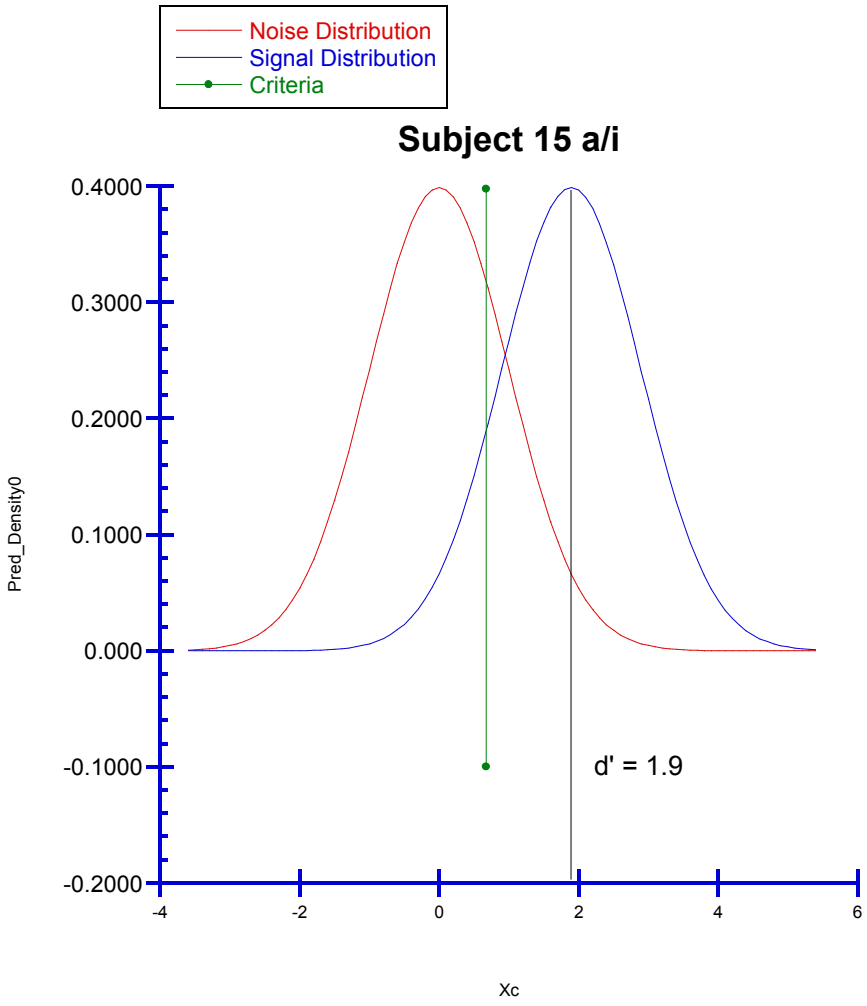


Figure 1. Subject 15's performance on the a/i contrast assessed via SDT.

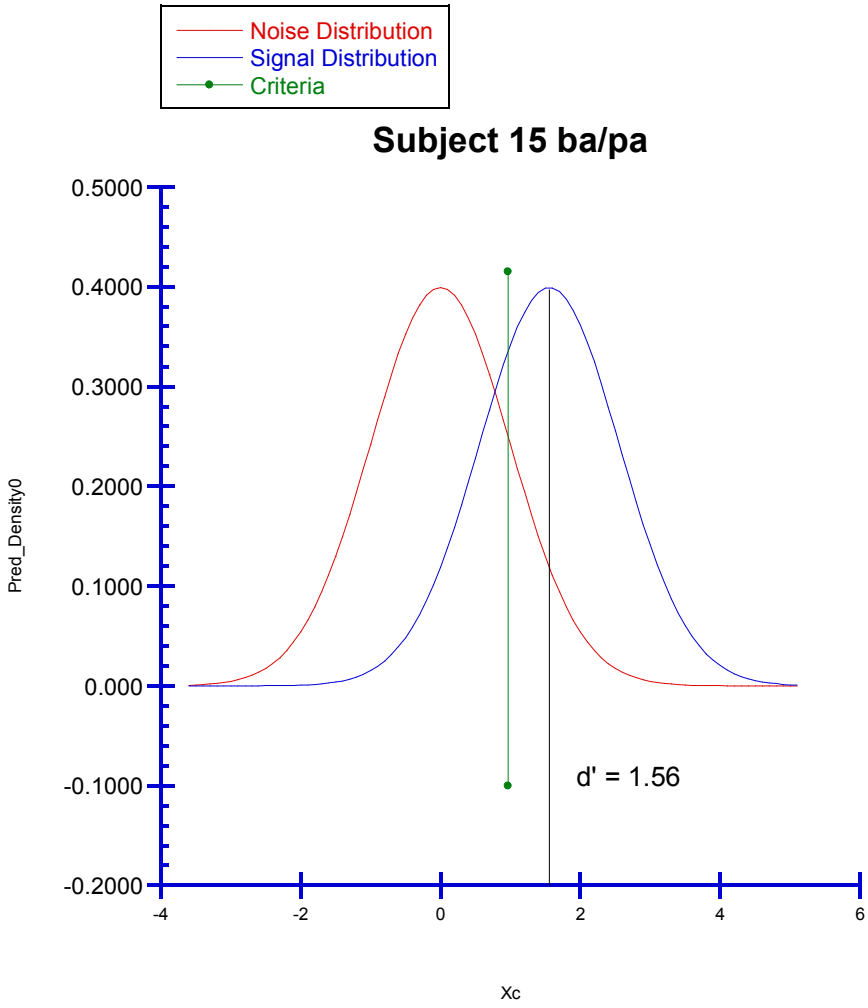


Figure 2. Subject 15's performance on the ba/pa contrast assessed via SDT.

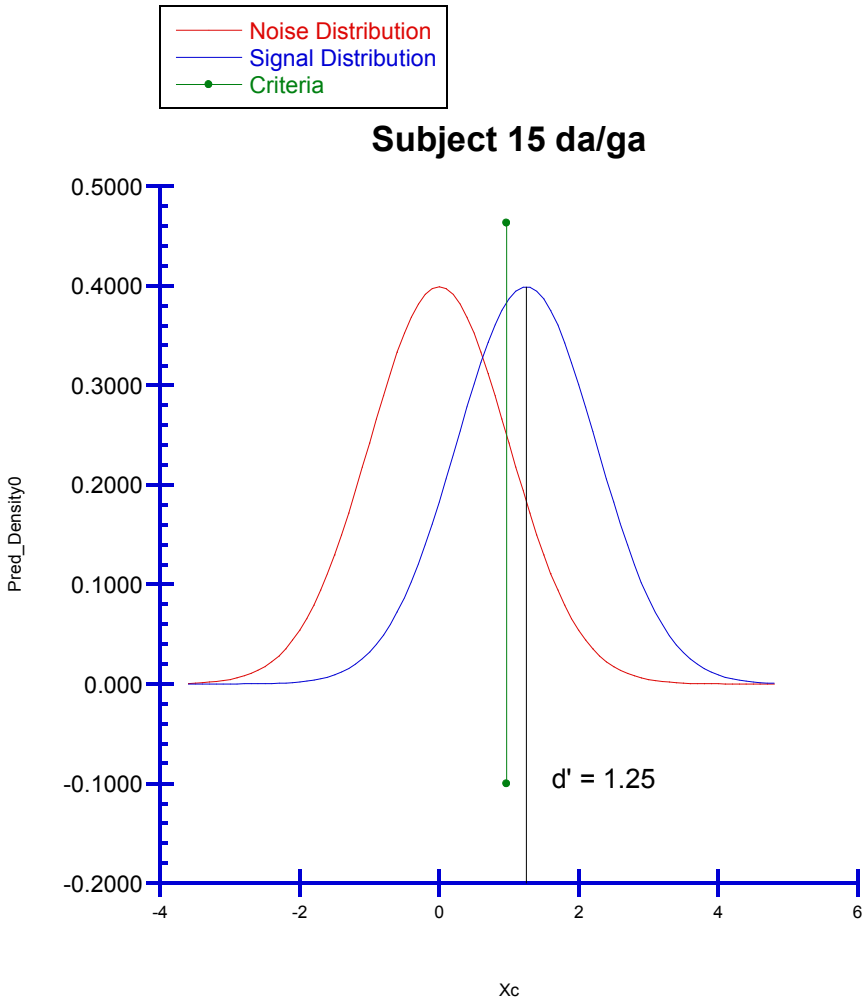


Figure 3. Subject 15's performance on the da/ga contrast assessed via SDT.

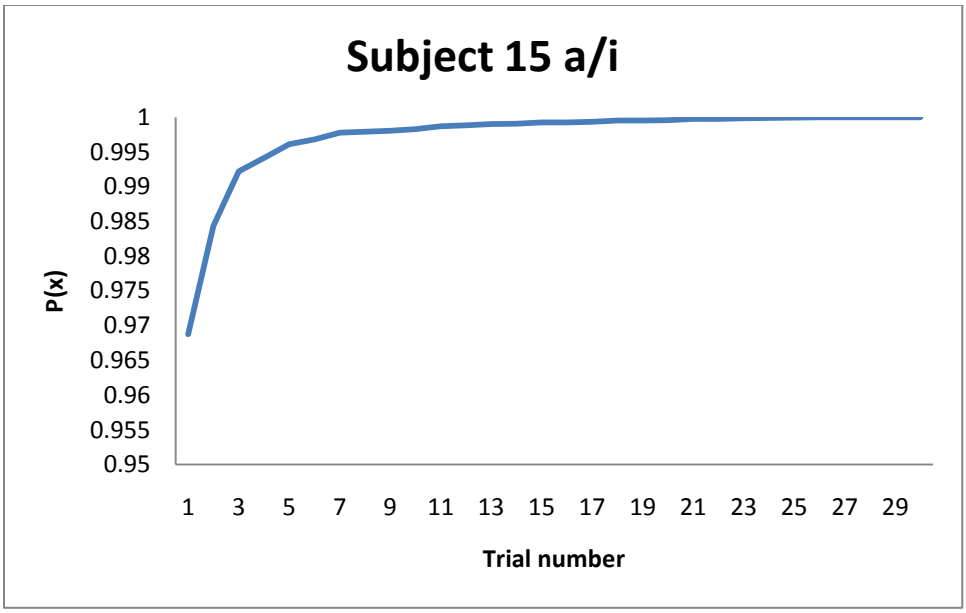


Figure 4. Cumulative distribution function for Subject 15's performance on the a/i contrast.

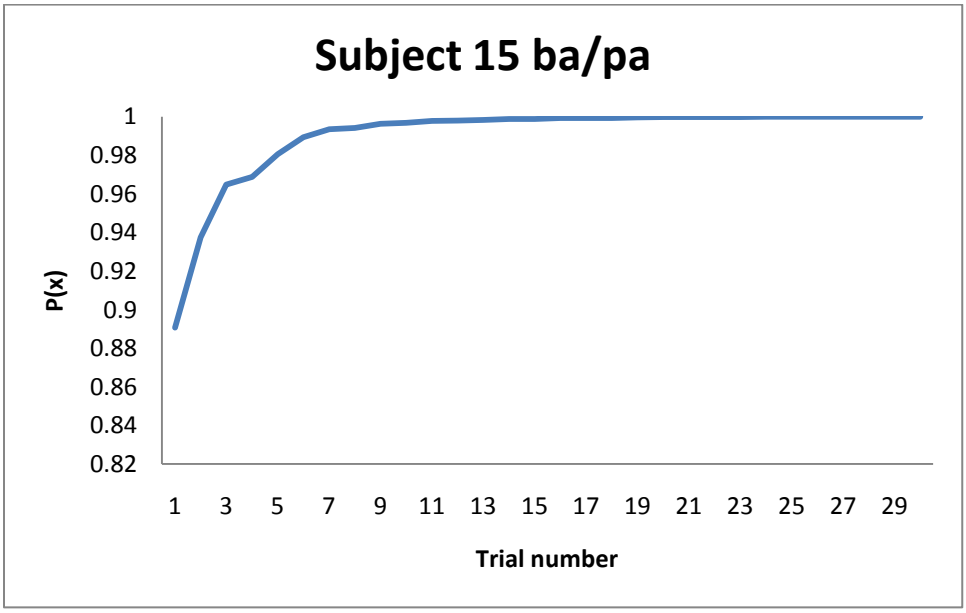


Figure 5. Cumulative distribution function for Subject 15's performance on the ba/pa contrast.

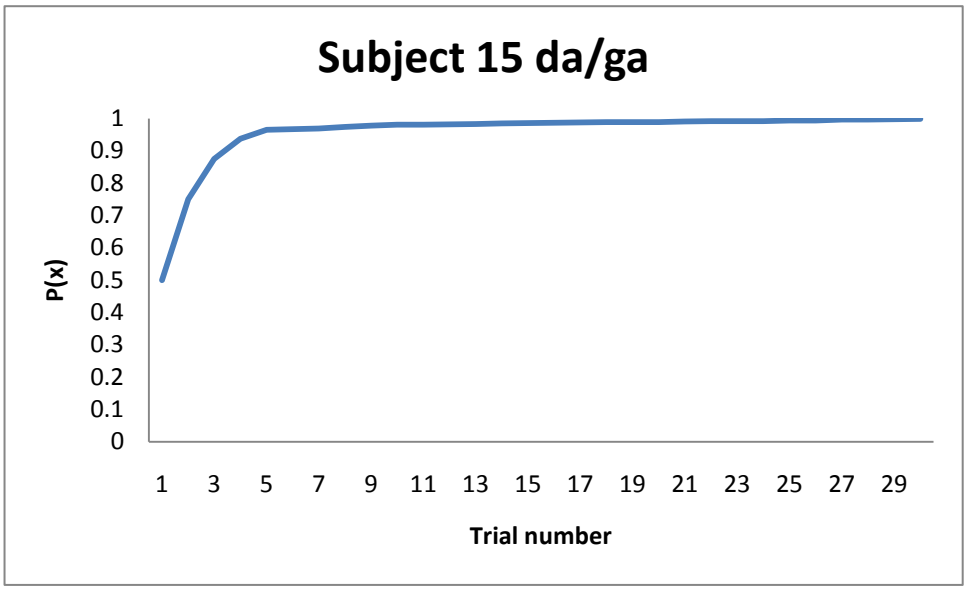


Figure 6. Cumulative distribution function for Subject 15's performance on the da/ga contrast.

Appendix P: Pseudo Subject responses

| | Subject | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trial | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| 2 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 |
| 3 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| 4 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| 5 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 6 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 7 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 8 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 9 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 10 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 11 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 |
| 12 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| 13 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 14 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 |
| 15 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 |
| 16 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| 17 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |
| 18 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| 19 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| 20 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 21 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| 22 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| 23 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| 24 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| 26 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| 27 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 28 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 |
| 29 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 30 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |

Table 1. Simulated trial-by-trial response data for 15 pseudo subjects. 1= correct response, 0 = incorrect response