6-2021

Infant Behavioral Speech Discrimination Procedures: A Systematic Review

Allison L. Mazzella
The Graduate Center, City University of New York

INFANT BEHAVIORAL SPEECH DISCRIMINATION PROCEDURES: A SYSTEMATIC REVIEW

By

ALLISON MAZZELLA

A capstone research project submitted to the Graduate Faculty in Audiology in partial fulfillment of the requirements for the degree of Doctor of Audiology, The City University of New York

2021

INFANT BEHAVIORAL SPEECH DISCRIMINATION PROCEDURES: A SYSTEMATIC REVIEW

By

ALLISON MAZZELLA

This manuscript has been read and accepted by the Graduate Faculty in Audiology in satisfaction of the capstone project requirement for the degree of Au.D.

________________________________ ________________________________

Date

Brett A. Martin, Ph.D., CCC-A
Faculty Mentor/Advisor

________________________________ ________________________________

Date

Brett A. Martin, Ph.D., CCC-A
Executive Officer

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Infant Behavioral Speech Discrimination Procedures: A Systematic Review

By: Allison Mazzella, Advisor: Brett A. Martin

Objective: The purpose of this capstone project was to conduct a systematic review of literature relating to two behavioral infant speech discrimination procedures to evaluate their potential clinical utility. The two procedures examined were the Observer-based Psychoacoustic Procedure (OPP) and the Visual Reinforcement of Infant Speech Discrimination (VRISD) method. The methodology utilized and the results obtained are examined for normal hearing infants and infants with hearing loss. The procedures are compared and contrasted in terms of potential clinical feasibility and modifications for potential clinical use are considered.

Methods: A comprehensive search was performed using Pubmed and EBSCO Academic Search Complete databases accessible through the City University of New York (CUNY) Graduate Center library to identify studies utilizing OPP or VRISD in infants under the age of 36 months. Inclusion criteria consisted subject age (under 36 months), use of speech stimuli or assessment of auditory abilities closely relating to speech perception, such as frequency and intensity discrimination, gap detection, and localization.

Results: Twenty-four studies met the inclusion criteria for this systematic review. Findings indicated that VRISD studies generally required fewer trials than OPP studies to obtain discrimination results; however, OPP allows for testing of younger infants. The amount of time required for testers to be adequately trained was longer for OPP compared to VRISD. Sample sizes differed for the 2 procedures. Both procedures controlled for observer and response bias; however, this was more consistent for OPP studies. Some differences in discrimination ability between infants with normal hearing compared to infants with hearing loss were noted, as would be expected. VRISD studies consistently used phonemic stimuli and produced repeatable results in infants with and without hearing loss. OPP studies utilized speech stimuli and other psychoacoustic stimuli and assessed abilities related to speech perception.

Discussion: While more research is warranted, both procedures have obtained repeatable results and an evaluation of the speech discrimination abilities of infants for a variety of speech contrasts can be performed. More studies employing larger sample sizes are needed for infants with hearing loss. VRISD may be easier to implement clinically as the procedures are more familiar to audiologists, requires less tester training, can be adapted to 1- and 2-tester conditions, and more quickly and easily provides discrimination results for different speech contrasts; however, OPP has an advantage in that it can be used with younger infants and the procedure incorporates more controls. Modifications must be made to procedures prior to clinical use, including standardization of stimuli and established shaping and training protocols. A proposed modification of the VRISD procedure is included for potential clinical use. These procedures, if used clinically, can potentially serve to monitor changes in speech discrimination with development, amplification and/or training.

Key words: infant speech perception, visual reinforcement, hearing loss in infancy, auditory development

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Brett Martin, for her guidance throughout my experience in this program. I would also like to thank Dr. Susan Wortsman and Dr. Dorothy DiToro for their guidance in the clinical aspects of this program. To my cohort, you have been the best support system throughout this entire process. I consider you all lifelong friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

As technology has improved, it has become commonplace to identify hearing loss shortly after birth. The average age of identification of congenital hearing loss in the United States is currently 3 months (Early Hearing Detection and Intervention Action Center [EHDI], 2017). More specifically, improvements in electrophysiological measurements such as otoacoustic emissions, auditory brainstem response, and auditory steady-state response testing have allowed professionals to estimate an infant's hearing thresholds relatively accurately. This has allowed for earlier fitting of hearing aids and cochlear implants.

The Joint Committee on Infant Hearing (JCIH) was formed in 1969 and composed of representatives from audiology, otolaryngology, pediatrics, and nursing with the goal to make recommendations related to the early identification of infant hearing loss and for newborn hearing screening (Joint Committee on Infant Hearing [JCIH], n.d.). In 2000, the JCIH created the Early Hearing Detection and Intervention program, which implemented universal newborn hearing screenings throughout the United States. The idea of 1-3-6, hearing screening before the age of 1 month, diagnosis before the age of 3 months, and intervention prior to the age of 6 months, was also implemented in the JCIH 2000 position statement, which is still relevant to clinical practice today. In the year 2000, the average age of identification of hearing loss was 2 years old (JCIH, 2000). In 2017, the Early Hearing Detection and Intervention Action Center (EHDI) reported that the prevalence of hearing loss was 1.7 per 1,000 children, with the average age of diagnosis being 3 months of age (EHDI, 2017). Follow-up to failed hearing screening and increased enrollment in early intervention has increased in the last decade. Per EHDI (2005), 64% (n=38,411) of infants that referred on hearing screenings were lost to follow up. Conversely, in 2018 25.9% of infants that referred on hearing screenings (n=15,581) were lost to follow-up. In 2005, 2,634 (0.9 per 1,000 screened) infants were identified with permanent hearing loss while in 2018 6,432 (1.7 per 1,000) infants were identified with permanent hearing loss (EHDI, 2005; EHDI, 2018). This likely reflects improvements in technology in the field, allowing for earlier identification of milder hearing losses.

Hearing thresholds in infants can be accurately estimated through behavioral and/or electrophysiological measurements. Examples of electrophysiological measurements include otoacoustic emissions (OAEs), tone-burst auditory brainstem response (ABR), and auditory steady-state response (ASSR). Per the American Academy of Audiology (AAA) pediatric clinical guidelines, the purpose of electrophysiologic measures in infants is "to determine [the] presence and type of hearing loss, and to estimate hearing levels for individual frequencies in each ear." (2020, p. 45). Testing is intended for infants, children, and those who cannot provide accurate hearing thresholds behaviorally. Testing requires the patient to be asleep, sedated, or calmly sitting. Sedation must occur in a controlled medical setting, which can be a caveat for some audiological practices.

As the infant reaches toddlerhood, electrophysiologic measures become increasingly more challenging. Toddlers are generally mobile, active, and do not tolerate electrophysiological testing as well as infants. Without the use of sedation, assessing toddlers through electrophysiologic measures may require multiple clinical visits and can yield minimal results on speech discrimination processes in children under the age of 2 years. Nonetheless, audiometric threshold can usually be estimated using behavioral audiometry. For speech perception testing, the Northwestern University Children's Perception of Speech (NU-CHIPS) test is widely used clinically and is normed for children aged 2-5 years (Auditec Incorporated, n.d.). Some measures, such as long-latency auditory evoked potentials, have proven to be useful in indexing the encoding (obligatory potentials) and discrimination (mismatch negativity, late positivity); however, these procedures can be costly, difficult to conduct in a typical clinical audiological setting, unavailable on some clinically available event-related potential devices, and the interpretation and analysis of these potentials requires more training than many audiologists receive. All of this means that while hearing loss can be identified well in young children, the ability to evaluate speech perception, which is needed to determine whether a child has access to the speech signal and whether the reception of speech is distorted needs work. The ability to evaluate speech perception capacity in these young children is also critically important for the fitting of hearing aids and cochlear implants to ensure that access to the speech signal is maximized.

It is evident that hearing loss manifested early in life can cause delays in psychosocial, communicative, and educational achievement (Eisenberg, Martinez, & Boothroyd, 2003, p. 327). Considering that auditory capacity is necessary for the acquisition of verbal language, a child with hearing loss may not develop speech and language without intervention. Eisenberg, Martinez, and Boothroyd (2007) describe three levels encompassing sound awareness and auditory perception in children. These levels include level I, which corresponds to sound awareness, level II, which corresponds to phonetic discrimination, and level III, which corresponds to word recognition (Eisenberg et al., 2007, p. 2). These levels infer that a child must be able to detect sounds and discriminate phonemes prior to successfully recognizing words. While behavioral measures are established for measuring hearing thresholds (level I) in infants, there is no established clinical procedure that measures phonetic discrimination (level II). While word recognition can be assessed in children around the age of 2 to 3 years old via the Northwestern University Children's Perception of Speech (NU-CHIPS), a child with hearing loss must be able to discriminate phonemes prior to successfully completing a word recognition task. Assessment of phonemic discrimination in these infants is warranted to track progression of auditory perception capabilities in this population.

Early intervention and active parental involvement, however, can potentially minimize the consequences of a childhood hearing loss. State requirements for early intervention vary, however a team approach including speech-language pathologists, audiologists, medical doctors, and other professionals are required to diagnose and treat a child's hearing loss. The approach with early intervention should be multi-disciplinary and parents or caregivers should be actively involved in the evaluation and habilitation processes.

Following the guidelines of the American Academy of Audiology (2013), children should be screened for hearing loss prior to one month of age, be diagnosed with hearing loss prior to three months of age and fit with amplification prior to six months of age to develop speech and language abilities in an adequate fashion (p. 5). The audiologist has a critical role in assuring that a child with hearing loss can hear speech at all sound levels and in a variety of environments. Therefore, the audiologist must accurately and efficiently verify and validate that the child's amplification technology is properly functioning. However, there are minimal behavioral testing measures that can be performed in those under the age of 3 years old. Prior to 2.5-3 years of age, the behavioral approach to testing utilizes visual reinforcement audiometry (VRA) to obtain minimum response levels to pure tone and speech stimuli. VRA is a procedure in which the child is conditioned to turn his or her head in response to the detection of a sound. A child responds by turning their head and the action is reinforced through the activation of a light up toy or video. The behavior (a head turn) is rewarded, thereby increasing the chances that the behavior will continue (Widen et al., 2002).

The American Academy of Audiology ([AAA], 2013) suggests the use of parental reports such as the Infant-Toddler Meaningful Auditory Integration Scale (IT-MAIS), LittlEars, or Parents Evaluation of Aural/oral Performance of children (PEACH) combined with aided pure tone testing and speech detection thresholds in the sound field (p. 41). Parent questionnaires and aided pure tone thresholds, however, do not provide an objective measure of speech discrimination ability.

These parent scales and questionnaires may be used to track progress and validate success with amplification; however, the clinician must acknowledge that the results of these questionnaires may not be reflect the child's speech perception performance in different listening situations or at different points in development. Moreover, questionnaires such as IT-MAIS and LittlEars do not assess the child's ability to discriminate speech sounds in minimal pairs such as differentiating /a/ from /i/ or /ba/ from /da/, which can aid in validation and programming. According to Uhler and colleagues (2018), "despite identification at earlier ages, there continue to be gaps in language outcomes between children with hearing loss and their peers with normal hearing" (p. 847). Since speech discrimination is the precursor to spoken language, communication, and literacy, the audiology community must attempt to objectively measure a child's speech discrimination ability in a sensitive and specific manner that is both valid and reliable.

Behavioral Approaches for the Evaluation of Speech Perception in Infants & Toddlers:

Although the clinical audiologic test battery lacks standard clinical measures of speech discrimination capacity for infants and toddlers, a variety of laboratory procedures have been successfully implemented such as Visual Reinforcement Infant Speech Discrimination (VRISD), Observer-based Psychoacoustic Procedure (OPP), Change/No Change, Visual Reinforcement

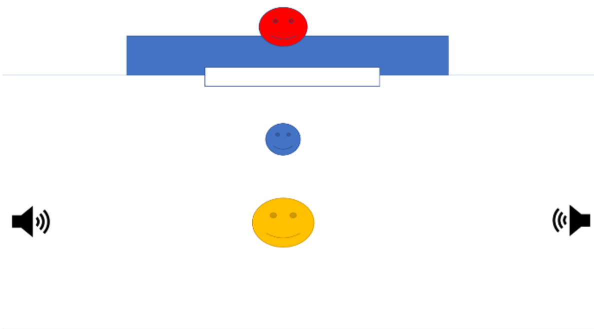
Assessment of Speech Pattern Contrast (VRASPAC), etc. These procedures have helped in determining the speech discrimination abilities of normal hearing children and are now being adapted for use in infants and toddlers with hearing loss. However, because these procedures are time-consuming, stringent, and require sophisticated instrumentation, they may not be practical for clinical application without substantial modification. This capstone research project will focus on the two procedures with the most literature: Visual Reinforcement Infant Speech Discrimination and the Observer-Based Psychoacoustic Procedure.

VRISD:

Visual Reinforcement Infant Speech Discrimination (VRISD) was developed in the 1970s to use conditioned head-turns to assess speech discrimination in infants aged 6-30 months (Moore, Wilson, and Thompson, 1977). In VRISD, the child is conditioned to turn to a visual reinforcer when they hear a change in stimuli. The child hears simple sounds such as /a/ or /i/ or consonant sounds such as /s/ and /sh/. One stimulus is the background stimulus and plays repeatedly while the other stimulus is the target stimulus, or the stimulus that will solicit a conditioned head turn. VRISD is a modification of VRA and is procedurally similar to VRA. Figure 1 presents a schematic of the set-up for VRISD. The infant is placed upon their caregiver's lap or in a highchair in a sound-treated room. The assistant and parent are in the booth with the infant. The assistant keeps the infant's attention to the midline and the experimenter is seated outside the booth monitoring the infant via video or through a window. When the infant appears attentive and ready, the experimenter initiates a trial. Instead of reinforcing the detection of a sound as in VRA, for VRISD the discrimination of a sound is reinforced. A full head turn is required to initiate reinforcement. VRISD protocol varies by researcher and institute. The child must be trained to turn his or her head in response to a change in stimulus. For example, the sound /ba/ might be presented repeatedly and occasionally replaced by /da/. A head turn in response to the /da/ stimulus would be reinforced.

A modification of VRISD, the visual reinforcement assessment of the perception of speech pattern contrasts (VRASPAC) was developed in 2004 by Eisenberg, Martinez, and Boothroyd. The procedure "combines the visual reinforcement infant speech discrimination (VRISD) test, the change/no change technique, and the speech pattern contrast concept" (Eisenberg et al., 2004, p. 365). VRASPAC was developed with the intention of investigating speech pattern contrast perception across ages groups and is part of a larger battery of tests based on the speech pattern contrast concept. For the purposes of this review, VRASPAC studies will be discussed and analyzed together with VRISD studies as the procedures and stimuli used are comparable.

Training procedures vary by research lab and will be discussed at length in the results section. The use of VRISD allows for the assessment of speech discrimination abilities independent of language abilities. VRISD was not used on children with hearing loss until after the year 2000 (Eisenberg et al., 2004; Martinez et al, 2008, Uhler et al., 2011; Uhler et al., 2018). Uhler and colleagues (2010) note that toddlers with hearing loss can demonstrate prelinguistic discrimination ability with amplification using a conditioned head turn.



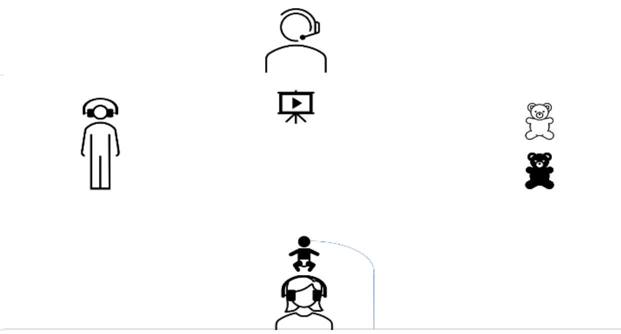


Figure 1(left): Sample setup for VRISD testing. The yellow face is the infant/caregiver, the blue face is the assistant, and the red face is the tester. The tester is operating equipment in a separate room, and the assistant and infant/caregiver are in a sound-treated room. Note that there may be differences in setup by research lab (i.e.: use of headphones, speakers at 45-degree azimuth, etc.)

Figure 2 (Right): Sample setup for OPP testing adapted from Lalonde and Werner 2019. The caregiver holds the child and listens to masking music. The assistant listens to the experimenter's instructions while the experimenter observes from a separate room via video. The infant is delivered auditory stimuli via an insert earphone placed in the right ear and visual reinforcers are placed to the right of the infant.

OPP:

The observer-based psychoacoustic procedure (OPP) combines features of the forced-choice preferential looking technique (Teller, 1979) and of visual reinforcement audiometry (Moore, Thompson, & Thompson, 1975). OPP responses are supposed to be objective and based on the observer's judgement as opposed to the infant's behavior. Per Olsho and colleagues (1987), "on every trial the observer is either right or wrong" regarding stimulus presentation parameters (p. 628). A forced choice procedure, such as OPP theoretically reduces the negative effects of response variability and the observer can base their response criteria on a variety of behaviors, "including direction of first look, duration of looking, head and body orientation, or any other cue" (Olsho, Koch, Halpin, & Carter, 1987, p. 628). A sample of OPP setup, adapted from Lalonde and Werner (2019) is shown in figure 2. In this experiment, the infant sits in the caregiver's lap inside a double-walled booth facing a TV monitor. An assistant is at the left of the monitor, manipulating quiet toys to keep the infant's attention toward the midline at the screen before the start of the trial. This experiment utilizes visual cues; in other experiments that utilize

OPP, there may not be a video monitor, however the general setup is like that of figure 2 in studies utilizing earphones.

Per Olsho and colleagues (1987), the OPP procedure can be described relative to either the infant or the observer; the observer's procedure "amounts to a yes-no task with feedback on each trial" (p. 628). The observer initiates a trial when the infant appears ready and an indicator light signals that a trial is in progress. The observer then watches the infant and much decide whether the signal (target sound) was presented. After each trial ends, the observer is informed whether a signal was presented. In terms of the infant being tested, the procedure is like that of a conditioned head-turn procedure, much like VRA or VRISD. However, the observer can report other apparent reactions as responses, in addition to a head turn, such as eye widening, eye movement, or change in arousal. The infant sits in a caregiver's lap "while listening to sounds presented monaurally over lightweight headphones" (p. 628). The infant's attention is guided towards the midline through an assistant manipulating toys. If the infant responds in such a way that the observer ***correctly*** decides a signal presentation has occurred, a mechanical toy is activated to reinforce whatever response is made by the infant. It is important to note that this response does not need to be a head-turn, and this is a key difference from VRISD. Correct observations, which are reinforced, are categorized as "hits" while incorrect observations, which are typically not reinforced, are labeled "false alarms". No feedback for negative responses or undetected positive responses is available to the infant during testing (Olsho et al., 1987, p. 629).

Olsho (1987), notes that training for the observer in OPP generally takes approximately one month (p. 631). The observer is considered "trained" when obtained thresholds are comparable to experienced observers and when s/he responds "yes" on no more than 25% of no-signal trials on a regular basis. The most common issue with training is a high false alarm rate, and therefore it is encouraged those observers adopt a “fairly conservative response bias” (Olsho et al., 1987, p. 631).

Results from VRISD or OPP Research:

Results from experiments utilizing VRISD or OPP have provided information on infant hearing thresholds, speech discrimination ability, localization ability, and gap detection abilities. For example, Hillenbrand and colleagues (1979) found that six- to seven-month-old infants could successfully discriminate /be-we/ contrasts and /be-ue/ contrasts utilizing the VRISD procedure. Results obtained reached statistical significance for these contrasts, however, results for another contrast, /we-ue/ did not achieve statistical significance. Moreover, Bull, Eilers, and Oller (1984) found that normally developing infants performed significantly better on two-syllable contrasts compared to three-syllable contrasts utilizing VRISD. Olsho and colleagues (1988) found that between the ages of 3 and 6 months, hearing thresholds improve in the higher frequencies while by the age of 12 months, threshold improvement can be seen in the lower frequencies (p. 1322). Werner and colleagues (1992) found that infants aged 3- and 6- months had significantly higher gap detection thresholds compared to adults and 12-month-old infants utilizing OPP. Trehub and colleagues (1991) utilized OPP on infants aged 1-4 months and found improvement in performance as a function of age.

More recent studies have utilized children with hearing loss with normal hearing children as a control group. Dasika and colleagues (2009) found that children with bilateral cochlear implants had significantly worse spatial acuity (localization) skills compared to children with normal hearing utilizing the OPP procedure. Moreover, they found that children utilizing bilateral cochlear implants performed significantly better in localization tasks compared to children with a unilateral cochlear implant (Dasika et al., 2009). These results are further corroborated by Greico-Calub and colleagues (2012) who found that infants with bilateral cochlear implants perform significantly better on localization tasks when using both implants rather than one implant alone. Results from Uhler and colleagues (2018) utilizing VRISD suggest that infants aged 7-28 months with hearing loss utilizing amplification struggle to discriminate consonant (ba-da) contrasts compared to infants with normal hearing. Additionally, they found that vowel contrasts are easier to discriminate than consonant. contrasts in both normal hearing children and infants with hearing loss (Uhler et al., 2018).

Rationale for research:

While results from behavioral speech discrimination procedures are promising for clinical practice, there are variations in procedural setup, stimuli, and training procedures that must be addressed prior to implementing these procedures in clinical practice and each approach has advantages and disadvantages.   Moreover, the typical behavioral speech discrimination procedure requires blinding, intensive calibration, and multiple testing sessions to conduct testing and obtain results. There are no established protocols for testing in the clinic, and it would be difficult to conduct testing utilizing an audiometer without access to universal pre-recorded and calibrated stimuli.

A universally available speech discrimination procedure that can be easily implemented in the clinic for infants is warranted for a variety of reasons. While input from parents on questionnaires such as PEACH or the IT-MAIS is helpful in tracking a patient’s progress, it is not a concrete measure of the patient's speech discrimination ability. Ideally, these questionnaires would be used to corroborate results obtained from audiological testing, rather than being used to track progress. Moreover, while electrophysiological testing can provide information on speech discrimination ability, many settings do not have the training necessary or have access to equipment that is required for testing the late potentials. Considering this, a behavioral speech discrimination testing procedure can be a valuable tool for audiologists working with infants with hearing loss. Assessment of discrimination ability with and without amplification can be useful for counseling purposes and tracking aural habilitation progress. Moreover, results can be used to fine-tune programming of devices.

It is important to assess speech discrimination abilities in children for a variety of reasons. First, it can provide an updated set of milestones in auditory development. To understand if auditory milestones are being met in a child with hearing loss, we must understand how a normal hearing child's auditory system typically develops. Second, audiologists can use speech discrimination tasks to measure a child's auditory development following intervention. These tasks can be administered repeatedly and used as an outcome measure and validation for hearing aid or cochlear implant fittings. Behavioral speech discrimination tasks can also aid the clinician's decision-making process in determining what interventions would be best for a child with hearing loss. For example, if an infant is unable to discriminate between specific speech sounds with amplification, more rigorous auditory training, or different intervention (e.g.: cochlear implantation) may be warranted. Third, this information can be used to evaluate progress with aural habilitation therapy.

The purpose of this capstone is to systematically review literature pertaining to these procedures and evaluate their efficacy and potential feasibility for use in a clinical setting. This review is warranted because behavioral speech discrimination procedures can be used to monitor changes in speech perception with development, with device fitting, and with aural habilitation. Should any of the mentioned procedures be used clinically, the audiologist would have substantial information pertaining to the child's speech discrimination and would be able to make gain adjustments accordingly as well as validate the child’s amplification. This may aid in optimal speech and language acquisition in child with hearing loss.

Goals:

1. Compare and contrast procedures
2. Evaluate potential clinical feasibility of each procedure
3. Compare application and results of each study for children with normal hearing and children with hearing loss
4. Consider modifications that would be needed for clinical implementation

METHODS

EBSCO Academic Search Complete and Pubmed databased were searched. Search filters included peer-reviewed journals with articles written in English and studies with participants aged 36-months-old and younger in the sample. The main search terms utilized were "visual reinforcement infant speech discrimination", "observer-based psychoacoustic procedure", "observer-based psychophysical procedure", "behavioral infant speech discrimination", "visual reinforcement assessment of speech pattern contrasts", "visual reinforcement infant speech discrimination", "infant speech discrimination", and "infant speech perception and psychoacoustics". The review utilizes the following inclusion criteria: articles published in English, subjects under the age of 36 months, and the use of the behavioral speech perception measures of Visual Reinforcement Infant Speech Discrimination (VRISD) and/or the Observer-based Psychoacoustic Procedure (OPP). Studies that assessed the perception of stimuli closely related to speech, such as frequency and/or loudness discrimination, or localization acuity were also included.

The results of the search are summarized in Figure 3. ESBCO searches yielded a total of 17 studies and Pubmed searches yielded a total of 177 studies; 11 duplicates were found, giving a total of 183 studies to be screened to meet inclusion criteria. 153 studies were excluded based on not meeting inclusion criteria, yielding 30 studies for in-depth assessment. Subsequently applied exclusion criteria were studies with children outside of the desired age range (n=2), studies that solely include event-related potential testing (n=2), and studies that did not utilize OPP or VRISD upon further examination (n=2), leaving 24 studies to be evaluated in this systematic review. Articles included in this review were assessed for type of behavioral testing procedure, shaping and training criteria, number of visits needed to complete testing, independent variables, dependent variables, sample size, stimuli, and results. Studies are also divided into subsections of the procedure used, observer-based psychoacoustic procedure (OPP) or visual reinforcement infant speech discrimination (VRISD).

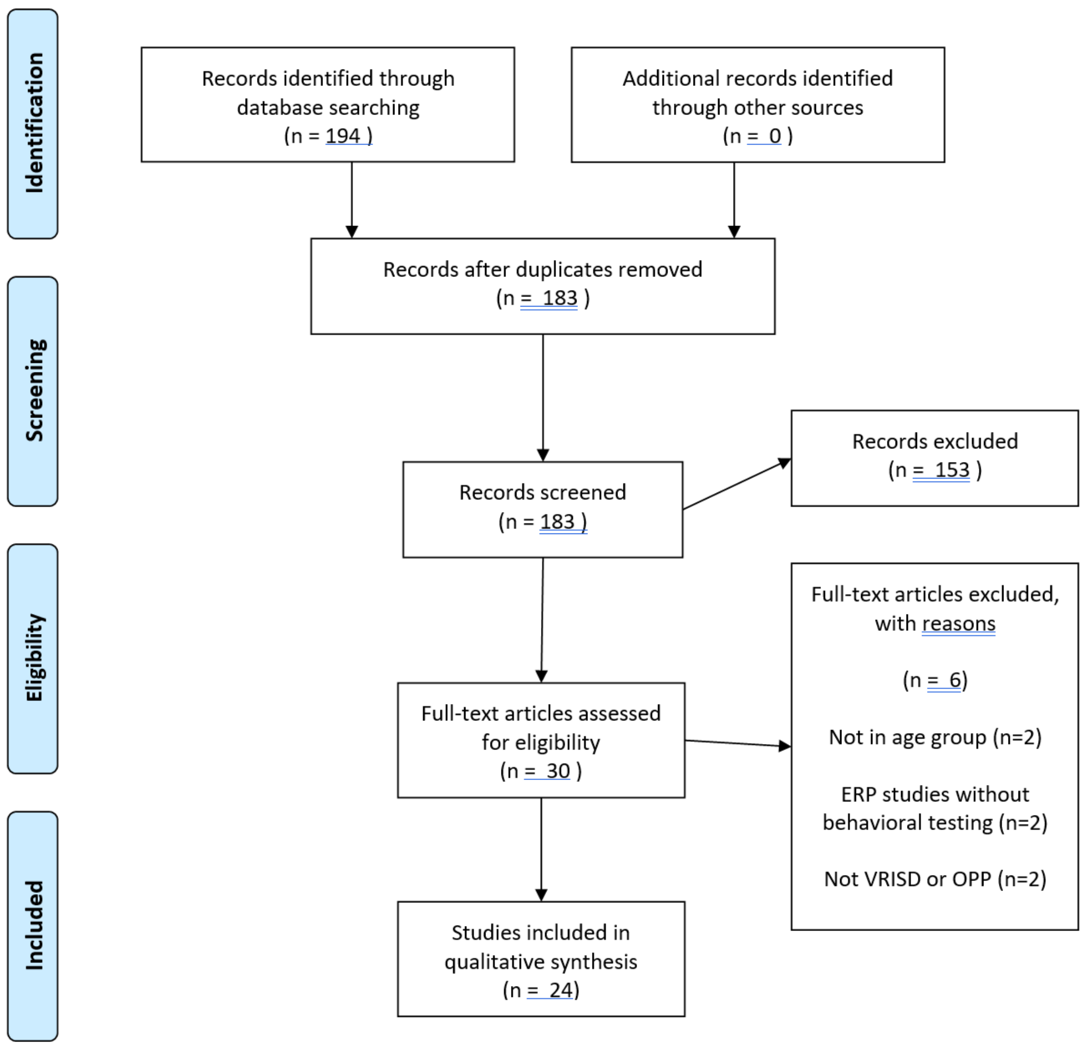


Figure 3: Flowchart of systematic review.

RESULTS

Of the 24 studies, 15 (62.5%) utilized Visual Reinforcement Infant Speech Discrimination (VRISD) or its variant Visual Reinforcement of Speech Pattern Contrasts (VRASPAC) and 9 (37.5%) utilized the Observer-based Psychoacoustic Procedure (OPP). The age of participants in VRISD studies ranged from 4 months of age to 38 months of age and the age of OPP participants ranged from 3 months of age to 36 months of age. 8 (5 VRISD, 3 OPP) studies included children with hearing loss while the remaining 18 utilized children with normal hearing. 26 studies were conducted in the sound field while 2 studies (both OPP) utilized an insert headphone in the right ear. None of the studies included ear-specific information for the left ear. Table 1 lists study design characteristics, including sample size, participants, inclusion/exclusion criteria, and statistical analyses used for each selected paper.

Table 1: Study Design Characteristics

Study	Procedure	Size	Participants	Inclusion/Exclusion criteria	Statistical Analysis
Eilers, Wilson, Moore 1977	VRISD	17	17 infants aged 1-3 months, 3-6 months, and 12 months with normal hearing	Infant considered "normal" if there were no concerns for development by parents, experimenters, and medical personnel and hearing was judged to be normal	Individual z-test for significance between 2 proportions
Hillenbrand, Minifie, Edwards, 1979	VRISD	9	9 infants aged 6-7 months with normal hearing	Parent questionnaire utilized to screen for children with history of ear infections, family history of hearing loss, and parental concern for infant's development	correlated t-tests
Eilers, Gavin, Wilson 1979	VRISD	10	Infants aged 6 months with normal hearing	No concerns from parents and medical personnel, normal hearing, presence of primary language spoken at home	2x2 mixed ANOVA with repeated measures
Goodsitt, Morse, Ver Hoeve, Cowan 1984	VRISD	16	Infants aged 6.5 months with normal hearing	None listed	Separate t-tests, mixed and redundant ANOVA
Hillenbrand 1984	VRISD	23	Infants aged 5.5-6.5 months with normal hearing	Parent questionnaire utilized to screen for children with history of ear infections, family history of hearing loss, and parental concern for infant's development	2 way ANOVA, 3 way ANOVA

Bull, Eilers, Oller 1984	VRISD	33	Infants aged 5-11 months with normal hearing	English speaking home, normal developmental milestones, normal hearing	z-test, ANOVA, Scheffe's post-hoc analysis
Sinnott, Aslin 1985	OPP	39	Infants aged 7-9 months with normal hearing and 6 adults with normal hearing	Infant considered in "good health", no colds or ear infection at time of visit, adults reported normal hearing	Arithmetic mean
Bull, Eilers, Oller 1985	VRISD	9	Infants aged 5-11 months with normal hearing	Normal developmental milestones, normal hearing	z-test, ANOVA
Nozza, Rossman, Bond 1991a	VRISD	23	15 infants age 9-11 months with normal hearing and 8 adults with normal hearing	Infants born between 38-42 weeks gestation, no neonatal risk factors for hearing loss or developmental delay. Adults: no significant history of otological disease or hearing loss	2 way ANOVA
Nozza, Miller, Rossman, Bond 1991b	VRISD	50	34 infants aged 7-11 months with normal hearing and 16 adults with normal hearing	Full term infants, free of parental concern regarding developmental delay or hearing loss, no reported cold or ear infection in the past two weeks. Adults: no significant history of otological disease or hearing loss	2 way ANOVA, absolute value of differences between 2 tests, t-test for differences between group means with unequal variances
Werner, Marean, Halpin, Benson, Spetner, Gillenwater 1991	OPP	50	39 infants aged 3-12 months with normal hearing and 11 adults aged 20-30 years with normal hearing	Infants born full term with normal pregnancy and post-natal development, no history of hearing dysfunction, no family history of hearing loss, no cold or middle ear infection within the past 2 weeks. Adults reported no hearing dysfunction	logistic regression, 3x4 ANOVA, one way ANOVA, arcsin transformation, Tukey post-hoc test

Eisenberg, Martinez, Boothroyd 2004	VRAPSAC	8	Eight children, 5 with normal hearing, 1 with mild hearing loss utilizing hearing aids, 2 with profound hearing loss utilizing bilateral cochlear implants aged 9-34 months	none listed	probability theory, confidence scores
Martinez, Eisenberg, Boothroyd, Visser-Dumont 2008	VRASPAC	20	11 infants with normal hearing, 9 with hearing loss ranging from mild to profound in severity. All children with HL utilized hearing aids, 4 children subsequently received cochlear implants aged 7-17 months	Normal hearing infants passed newborn infant hearing screenings bilaterally, and passed OAE and tympanometry screenings bilaterally	confidence scores
Grieco-Calub, Litovsky, Werner 2008	OPP	26	18 infants with severe-profound hearing loss, 10 utilizing bilateral cochlear implants, 8 using unilateral cochlear implants, 8 normal hearing infants aged 26-36 months	Infants with hearing loss: severe to profound hearing loss with unilateral or bilateral cochlear implant (CI) Normal hearing infants: no reported difficulies from parents, no middle ear issues, no developmental delay	logistic regression
Dasika, Werner, Norton,	OPP	12	12 cochlear implant recipients, 11 unilaterally	Infants with severe to profound hearing loss treated by at least one cochlear implant	p(C)max, d', confidence interval

Nie, Rubinstein 2009			implanted 1 bilaterally implanted aged 14-32 months		
Uhler, Yoshiaga-Itano, Gabbard, Rothpletz, Jenkins 2011	VRASPAC	10	3 infants with severe to profound hearing loss who utilized hearing aids and subsequently received cochlear implants aged 12-16 months at the initiation of study and 7 infants with normal hearing	Middle ear status screened for all groups. Developmental delay ruled out via Kent Infant Development Scale (Katoff et al., 1978), Minnesota Child Development Inventory (Ireton, 1992), and/or Words and Gesture Inventory of the MacArthur Communication Development Inventories (Fenson et al., 1993; Thal et al., 2007)	95% confidence interval
Grieco-Calub and Litovsky 2012	OPP	47	27 bilateral cochlear implant users, 12 unilateral cochlear implant users, 8 normal hearing infants aged 26-36 months	Infants with hearing loss: severe to profound hearing loss with unilateral or bilateral cochlear implant (CI) Normal hearing infants: no reported difficulties from parents, no middle ear issues, no developmental delay	arithmetic mean, paired t-test multivariate linear regression analyses
Cone and Whittaker 2013	OPP	45	36 infants aged 4-12 months and 9 adults, all with normal hearing	Infants born full term, passed newborn infant hearing screening, passed DPOAE and tympanometry screenings before testing. Adults passed a pure tone screening and reported to history of otologic disease	ANOVA, t-tests, d'

Cone 2015	VRISD	20	infants aged 4-11.8 months with normal hearing	All infants born full-term, no risk factors for hearing loss, passed newborn infant hearing screening, otoscopy, tympanometry, and DPOAE screenings passed before admission to study	ANOVA
Uhler, Baca, Dudas, Fredrickson 2015	VRISD	22	Infants aged 6-14 months with normal hearing	Pure tone testing 500-4kHz, speech awareness testing, tympanometry all within normal limits to be included in the study. Test of Auditory Skill Development (Meinzen-Derr et al., 2007) administered to all subjects. Subjects must not have developmental delay, subjects must be able to complete a head turn for testing, subjects must have a normal tympanogram the day of testing, subjects must come from an English or Spanish speaking household	Kaplan-Meier estimation procedure, logrank test, logistic regression generalized estimating procedure
Leibold , Bonino, Buss 2016	OPP	23	7 infants aged 8.3-10.1 months with normal hearing, 10 school-aged (8-10 years) with normal hearing,8 adults with normal hearing	No risk factors by self/parent report, no treatment for Otitis Media for at least 1 week, no more than 2 years of musical training	power analysis, arithmetic mean, repeated measures ANOVA

Uhler, Gifford, Forster, Tierney, Claycomb, Werner 2018	VRISD	43	21 infants with normal hearing, 22 infants with bilateral hearing loss. Of included infants with hearing loss (n=20), 17 used bilateral hearing aids and 3 used bilateral cochlear implants	Subjects must come from an English or Spanish speaking home, subjects have no reported developmental delay, subjects must be able to turn their head for testing, subjects must have normal tympanograms or patent pressure equalizing tubes the day of testing, daily use of hearing aids or CI as reported by parents, subjects with auditory neuropathy spectrum disorder excluded	p(C)max, d', confidence interval
Lalonde, Werner 2019	OPP	118	88 infants aged 6-8.5 months with normal hearing and 30 adults with normal hearing aged 18-30 years	All adults reported normal hearing and were native English speakers. All infants passed newborn infant hearing screening, had no family history of hearing loss and no risk factors for hearing loss. All subjects passed tympanometry screening before testing	d', linear mixed modelling
Walker, Gerhards, Werner, Horn 2019	OPP	49	33 infants aged 3 months with normal hearing, 16 adults with normal hearing	All infants born full-term, no risk factors for hearing loss, passed newborn infant hearing screening, no diagnosis of hearing loss or developmental delay, no otitis media episodes within 3 weeks of testing, no more than 2 otitis media episodes	arithmetic mean, two-tailed t-test for independent samples, p(C)

Table 2 displays variables and outcome measures by each procedure. Procedural setup, including which procedure used in each paper, is described in detail.

Table 2: Study Variables and Procedures

Study	Independent Variable(s)	Outcome measures	Procedure
Eilers, Wilson, Moore 1977	Stimulus contrasts: /sa-va/, /sa-sha., /sa-za/, /as-a:z/, /a:s-a:z/, /at-a:d/, /a:t-a:d/ /at-a:t/, /fa-tha/, /fi-thi/, age group: 1-3 months, 6-8 months, 12 months	45-degree head turn in response to a change in stimulus	VRISD: Infant sat in caregiver's lap in testing booth. Experimenter 1 kept infant's attention towards the midline with toys. Once infant is attending to experimenter 1, background stimulus is presented 1 syllable per second at 50 dB SPL. When the infant appeared ready, the target stimulus was introduced
Hillenbrand, Minifie, Edwards, 1979	Stimulus contrasts: /be-we/, /be-ue/, /we-ue/ both natural and synthetic	45-degree head turn in response to a change in stimulus	VRISD: all trials began when infant was in a "ready state". Experimenter initiated a 4 second observation interval with a 0.5 probability of a change trial occurring, 2 experimenters must vote "yes" to a head turn response for reinforcer to be activated

Eilers, Gavin, Wilson 1979	Differences in voice-onset time (VOT): +70 vs +40, +40 vs +10, +10 versus -20, -20 versus -50, 0 vs -30, +10 versus -60 lag, English versus Spanish language learners	Infant is conditioned to turn his head when he detects a change from stimulus 1 to stimulus 2 to receive visual reinforcement	VRISD: 3 change intervals and 3 control intervals presented to each child so that 6 discrimination scores obtained for each stimulus pair
Goodsitt, Morse, Ver Hoeve, Cowan 1984	Stimuli: /ba/ /du/ /ko/ /ti/ presented in various bisyllabic versus trisyllabic contexts; redundant versus non-redundant conditions, redundant conditions have two of the same syllable with one target syllable embedded while non-redundant contexts have different syllable with target syllable embedded	Head turn in response to change in stimulus	VRISD: Testing began approximately 1 week following training. The infant was briefly re-trained with the familiar training contrast. The background stimulus continuously repeated at 54 dB. An assistant distracts the infant with toys. Once infant appears ready, the experimenter initiates testing. Experimenter delivered reinforcement when the infant correctly turned their head in response to a change in stimulus
Hillenbrand 1984	1. /ma-mi-mu/ vs /na-ni-nu/ discrimination ability by both 2. male and female speakers	Head turn in response to change in stimulus	VRISD: Tape recorded stimuli continuously presented in test booth with an onset-offset rate of 1.7 seconds. Assistant keeps infants attention by playing with toys. When assistant judged infant to be in a ready state he pressed a button to signal the experimenter to initiate a 5 second observation interval. A hand-held vibrotactile device signaled the assistant of the start of a change trial while a small light mounted on the observation monitor signaled the experimenter of a change trial. If both experimenter and assistant agree that a head turn response has taken place, reinforcer is activated for 3 seconds

Bull, Eilers, Oller 1984	1. Discrimination of bisyllabic versus trisyllabic stimuli: /samad/ and /masamad/ and 2. stimulus change in intensity of +30, +20, +10 for final syllable	30-degree head turn in response to change in stimulus	VRISD: Assistant keeps infants attention by waving toys. At the beginning of a session, background stimulus presented at 60 dBC. A trial was initiated when the infant was facing the experimenter, at which experimenter pressed a button to initiate a 6 second observation period. Test phase consisted of 30 trials, approximately half of which were control trials. After every 5 trials, a probe trial was presented with a target stimulus presented 5 dB above background stimulus. Head turns during probe trials were not included in analyses.
Sinnott, Aslin 1985	1. Intensity and 2. frequency discrimination ability to a modulated 1kHz tone	Head turn response towards reinforcer in response to change in stimulus	OPP: Experimenter watches infant via video camera in a separate room. The experimenter operated a response box to code the infant's head turns. Throughout testing, a constantly repeating 1kHz tone was pulsed at a rate of 1 per 750 ms. To minimize patterned presentation of trials, each trial sequence began with a variable duration (9-14s) during which the experimenter could not initiate trials because their response button was inactive.
Bull, Eilers, Oller 1985	1. Discrimination ability of bisyllabic and trisyllabic stimuli: /samad/ and /masamad/ 2. ability to detect stimulus change in fundamental frequency peak (increase) of +30, +20, +10 for final syllable	30 degree head turn in response to change in stimulus	VRISD: Assistant keeps infants attention by waving toys. At the beginning of a session, background stimulus presented at 60 dBC. A trial was initiated when the infant was facing the experimenter, at which experimenter pressed a button to initiate a 6 second observation period. Test phase consisted of 30 trials, approximately half of which were control trials. After every 5 trials, a probe trial was presented with a target stimulus presented 5 dB above background stimulus. Head turns during probe trials were not included in analyses.

Nozza, Rossman, Bond 1991a	/ba-da/ versus /ba-ga/ discrimination ability	Head turns in response to a change in stimulus	VRISD: Assistant used mild levels of activity with small toys to keep the infant's attention. When the infant was focused on the midline, the experimented pressed a key on a computer to initiate the trial. Upon head turns during trials, the experimenter pressed a response button interfaced with the computer. Infant head turns following changes in speech stimuli were visually reinforced.
Nozza, Miller, Rossman, Bond 1991b	Ability to discriminate /ba-ga/ contrasts in various signal-to-noise ratios	Head turn in response to change in stimulus	VRISD: Repeating background stimulus presented from loudspeaker ever 1500 ms as infant entered the test room. Assistant used mild levels of activity with small toys to keep the infant's attention. When the infant was focused on the midline, the experimented pressed a key on a computer to initiate the trial. Upon head turns during trials, the experimenter pressed a response button interfaced with the computer. Infant head turns following changes in speech stimuli were visually reinforced. Changes in stimulus intensity changed the signal to noise ratio as masking noise was kept at the same level throughout testing.

Werner, Marean, Halpin, Benson, Spetner, Gillenwater 1991	Gap detection ability	Observer judgement of whether a signal trial had occurred	OPP: Observer watched infant from a separate room, either via video or through the window on the other side of the test booth. The observer began a trial when the infant was quiet and attending towards the midline. Noise stimulus was presented continuous throughout sessions. Signal trials were presented to the infant with a probability of 0.65. A flashing LED light indicated to the observer that a trial was taking place, however the observer was blinded to the type of trial (control versus signal). The observer used the infant's behavior to judge whether a signal trial had occurred.
Eisenberg, Martinez, Boothroyd 2004	Discrimination of stimuli: "doo", "daa", "dee", "too", "zoo", "boo", "goo" with "doo" serving as background stimulus	Head turns in response to a change in stimulus	VRASPAC: Child is seated on caregiver's lap and a test assistant in the booth maintains the child's attention. Standard syllable "doo" is repeated continuously until the child habituates to the standard. The contrast syllable (ie: "daa") is then presented and the child is conditioned to turn toward the reinforcer.
Martinez, Eisenberg, Boothroyd, Visser-Dumont 2008	Discrimination of vowel-consonant-vowel (VCV) utterances, including both vowel and consonant contrasts of /oodoo/-/aadaa/, /oodoo/-/eedee/, /oodoo/-/ootoo/, /oodoo/-/oozoo/, /oodoo/-/ooboo/, /oodoo/-/oogoo/	Head turn in response to change in stimulus	VRASPAC: Infant sits in caregiver's lap while caregiver listens to music through headphones. String of VCV utterances with the standard being "oodoo" is presented through the loudspeaker. When a phonetic contrast is introduced, ie: aadaa, the child is conditioned to turn towards the reinforcer.

Grieco-Calub, Litovsky, Werner 2008	1. Localization ability in normal hearing infants versus unilateral CI using infants versus bilaterally implanted CI users 2. Fixed versus roving conditions	Observer judgement of where a signal trial had occurred	OPP: right-left discrimination task. Observer is unaware of stimulus location. The observer signaled the computer to initiate a stimulus trial randomly presented to either side. After stimulus presentation, the observer made a judgement regarding stimulus location based on the infant's response
Dasika, Werner, Norton, Nie, Rubinstein 2009	1. Psychometric function for detection 2. Reaction time	Observer judgement of whether a signal trial had occurred	OPP: Five signal trials were presented at varying levels within a block of 12 total trials in attempt to encompass the psychometric function for detection. To maintain the child's interest in testing, reinforcement was varied between an animated toy and a video. If the child appeared to lose interest, the observer could insert a "probe" trial, on which the signal was presented at comfort level established by cochlear implant programming software. Breaks were taken approximately every 10-15 minutes or as needed
Uhler, Yoshiaga-Itano, Gabbard, Rothpletz, Jenkins 2011	stimulus contrast: /a-u/, /a-i/, /u-i/, /ta-da/, /pa-ka/	Head turn in response to change in stimulus	VRASPAC: Subject placed in caregiver's lap in a sound-treated room. Test assistant is seated in front of the child to maintain their attention. Both the caregiver and test assistant were masked. When the child was initially brought into the room, the background stimulus was presented continuously. When the child appeared ready, the experimenter initiated a trial. Only correct responses are reinforced.

Grieco-Calub and Litovsky 2012	1. Spatial acuity in normal hearing infants versus infants with unilateral CI versus infants with bilateral CI 2. Fixed versus roving conditions	Observer judgement of where a signal trial had occurred	OPP: Children's CI processors were programmed by their audiologists prior to the first visit. Right-left discrimination task. Observer is unaware of stimulus location. The observer signaled the computer to initiate a stimulus trial randomly presented to either side. After stimulus presentation, the observer made a judgement regarding stimulus location based on the infant's response
Cone and Whittaker 2013	1. Discrimination of4 tone bursts (0.5, 1.0, 2.0, 4.0 kHz), 2. discrimination of 7 speech tokens (i, a, o, u, m, sh, s) 3. Infant versus adult results 4. CAEP results versus behavioral results	Observer judgement of whether a signal trial had occurred.	OPP: Infants seated in parent's lap or a high chair. Test assistant manipulated toys to keep the infant quiet and alert. Parent and test assistant were masked with music over headphones. The observer initiated a trial when the infant appeared ready, however the observer was blinded to whether or not a signal trial had occurred. No reinforcement for misses, correct rejections, or false alarms
Cone 2015	1. Discrimination of vowel tokens /a/, /i/, /o/, and /u/.2. CAEP results versus behavioral results	Head turn, change in eye movement, or change in arousal in response to change in stimulus	VRISD: Infants brought into test booth while control stimulus is continually playing. A test assistant keeps the infant attentive. Testing phase consisted of 12 trials of a given vowel contrast. The observer initiated trials, however they were blinded to whether it was a change or control trial. Percentage correct of change trials and control trials were calculated.

Uhler, Baca, Dudas, Fredrickson 2015	1. /a-i/ discrimination ability, 2. /ba-da/ discrimination ability 3. /a-i/ versus /ba-da/ discrimination ability	Head turn in response to change in stimulus	VRISD: Infant seated in caregiver's lap on in a high chair. Background stimulus was playing as the infant entered the room. An assistant kept the infant attentive. The evaluator was seated on the opposite side of the test booth, observing the infant through a window. The evaluator initiated trials once the infant's attention was at the midline. VRISD software allowed for only correct head turn responses for changes in stimulus to be reinforced. The evaluator was blinded to whether a signal trial occurred.
Leibold , Bonino, Buss 2016	1. Infant ability to discriminate sounds in two-talker babble versus speech-shaped noise 2. Infant performance compared to school-aged children and adults	Observer judgement of whether a signal trial had occurred.	OPP: Target words were counterbalanced among listeners. Thresholds for target word were measured adaptively using a two-down, one-up procedure. The probability of a signal trial was 0.75, the probability of a catch trial was 0.125, and the probability of a probe trial was 0.125. Probe trials were utilized to re-familiarize infants with the procedure. Eight reversals were obtained and the mean threshold level was calculated as the mean signal level of the last 6 reversals. Procedural setup was the same across all groups. School-aged children and adults completed testing without an assistant in the room and responded by raising their hand.

Uhler, Gifford, Forster, Tierney, Claycomb, Werner 2018	1. /a-i/ discrimination ability, 2. /ba-da/ discrimination ability 3. /a-i/ versus /ba-da/ discrimination ability 4. differences in results between normal hearing and infants with hearing loss.	Head turn in response to change in stimulus	VRISD: Infant seated in caregiver's lap on in a high chair. Background stimulus was playing as the infant entered the room. An assistant kept the infant attentive. The evaluator was seated on the opposite side of the test booth, observing the infant through a window. The evaluator initiated trials once the infant's attention was at the midline. VRISD software allowed for only correct head turn responses for changes in stimulus to be reinforced. The evaluator was blinded to whether a signal trial occurred. Infants with hearing loss utilized amplification during testing.
Lalonde, Werner 2019	1. Detection ability in infants with auditory only versus auditory visual cues 2. Detection ability in adults with auditory only versus auditory visual cues. 3. Discrimination ability in infants with auditory only versus auditory-visual cues. 4. Discrimination ability in adults with auditory only versus auditory visual cues.	Observer judgement of whether a signal trial had occurred.	OPP: Participants trained to respond to a signal. No-signal trials were randomly presented throughout testing. Observers were blinded to whether a signal trial had occurred. Participants testing in 3 conditions: auditory-only, auditory visual, and onset-offset cue.

Walker, Gerhards, Werner, Horn 2019	1. Modulation detection thresholds in infants 2. Modulation detection thresholds in adults 3. Temporal modulation cutoff frequencies in infants 4. Temporal modulation cutoff frequencies in adults	Observer judgement of whether a signal trial had occurred	OPP: Listener seated in a chair or in their caregiver's lap. For infants, an assistant sat in the booth manipulating toys to maintain the infant's attention. The caregiver, assistant, and observer were masked via headphones. Listener was presented with repeating stimuli with a period of 2 seconds between each stimulus. The type of trial (control vs. signal) was presented randomly and the observer had a 4 second window to judge whether a signal trial had occurred.

Table 3 displays significant results found in each selected OPP paper. Results from studies will be discussed in depth in the “Results Obtained: OPP” portion of this review.

Table 3: Significant results in OPP studies

Study	Significant Results
Werner, Marean, Halpin, Benson, Spetner, Gillenwater 1991	Significant effect for age (p<0.001) and masking condition (p<0.001) effect of age is significant for 500 Hz, 2000 Hz, and 8000 Hz masker cutoff conditions (p<0.001) and masking conditions (p<0.001)
Dasika, Werner, Norton, Nie, Rubinstein 2009	Mean reaction time for false alarms on no-signal trials was significantly greater than for hits on signal trials (p<0.001).
Grieco-Calub and Litovsky 2012	Significant difference in minimal audible angle (MAA) in infants using BICI in the fixed condition compared to normal hearing infants in the fixed condition (p<0.001). BICI group performed significantly better than UCI group (p<0.001). Children with BICI performed significantly better using both processors compared to one alone (p<0.001)

Cone and Whittaker 2013	Significant differences in threshold for each subject group (p<0.0001). Significant differences between infant and adult performance (0.0001). Adult thresholds for /u/, /m/, and /s/ significantly elevated compared to thresholds for /a/, /i/, /o/, and /sh/ thresholds (p<0.0001).
Leibold , Bonino, Buss 2016	Significant effect for masker type (p<0.01), age group (p<0.001), and masker type x age group (p<0.01). Response bias testing revealed a main effect for age (0.001)
Lalonde, Werner 2019	Detection: significant differences between auditory only condition and onset-offset cue condition across groups (p=0.0127). Discrimination: significant difference between auditory-only and auditory visual conditions (p=0.0305). Significant difference between auditory only and onset-offset cue conditions (p=0.0031). Significant interaction for onset offset ersus auditory only condition in infants (p=0.0066). Significant differences between AV cue and onset-offset condition with /lu/ target across groups (p=0.0351).
Walker, Gerhards, Werner, Horn 2019	Significant differences between infants' and adults' modulation detection thresholds (p<0.0001)

Table 4 displays significant results obtained in VRISD papers. Results from each paper will be discussed in depth in the "Results Obtained: VRISD" section of this review.

Table 4: Significant Results in VRISD Studies

Study	Significant Results
Eilers, Wilson, Moore 1977	Significant evidence of discrimination (p<0.05) for /va-sa/, /sa-sha/, /as-a:z/, /at-a:d/ in infants aged 1-3 months. Results also significant for those stimuli plus /sa-za/ in 6 month-olds (p<0.05)
Hillenbrand, Minifie, Edwards, 1979	Significant evidence of discrimination (p<0.05) for /be-we/ and /be-ue/ contrasts using synthetic speech. Significant evidence of discrimination for /be-we/ (p<0.05), /be-ue/ (p<0.005), and /we-ue/ (p<0.005) with natural speech
Eilers, Gavin, Wilson 1979	Significant main effect for stimulus pair (p<0.01) and significant language experience x stimulus pair interaction (p<0.01)
Goodsitt, Morse, Ver Hoeve, Cowan 1984	Performance of both target groups was above chance (p<0.01)
Hillenbrand 1984	Significant main effect for trial type (p<0.005) and trial type effect is significant for the phonetic group (p<0.05)
Bull, Eilers, Oller 1984	Discrimination index (DI) scores significantly above chance performance. Significant effects for increment (p<0.001) and syllable number (p<0.01). Discrimination for 2dB increments differs significantly compared to 4dB increments (p<0.05) or 6 dB (p<0.01)

Bull, Eilers, Oller 1985	All discrimination index scores significantly above chance performance for all contrasts (p<0.01). ANOVA comparing infant and adult performance on 3 contrasts yielded significant main effects for age group (p<0.01) and frequency increment (p<0.01). Significant interaction between main effects noted (p<0.01 for +10- and +30-Hz increments
Nozza, Rossman, Bond 1991a	2-way ANOVA with repeated measures revealed main effects of age (p<0.001) and contrast (P<0.01)
Nozza, Miller, Rossman, Bond 1991b	Differences between average SNR required to discriminate sounds was significantly different in infants versus adults (p<0.01)
Cone 2015	ANOVA of hit rate as a function of contrast type was significant (p<0.05)
Uhler, Baca, Dudas, Fredrickson 2015	Significant difference between stimulus contrast and criterion (p=0.03). Significant difference between groups who reached criterion and groups who did not reach criterion for /ba-da/ in loudest sensation level used (p=0.004). Significant effect for contrast (p=0.009). Results suggest that contrast (p=0.001) and sensation level (p=0.009) are significant predictors of criterion
Uhler, Gifford, Forster, Tierney, Claycomb, Werner 2018	For infants with hearing loss, aided speech intelligibility index (SII) and high-frequency puretone average were significantly related to success on /a-i/ contrast.

Stimuli:

Table 2 shows stimuli used in each selected paper. Stimuli used in the selected studies vary depending on the intended research. Stimuli in OPP studies included gaps in noise (Werner et al., 1991), and Gaussian noise (Walker et al., 2019). Some studies used multisyllabic stimuli, such as the spondaic words "baseball" and "birthday" in a localization task (Greico-Calub, Litkovsky, & Werner, 2008; Greico-Calub & Litkovsky, 2012) or disyllabic words such as "baby", "tiger", and "ice cream" in a speech detection in noise task (Leibold, Bonino, & Buss, 2016). Cone (2015) utilized /a-i/, /a-o/, and /a-u/ contrasts, while Lalonde and Werner (2019) utilized consonant-vowel stimuli including /mu-gu/, and /mu-lu/ contrasts.

Studies utilizing VRISD, including studies utilizing VRASPAC, also used a variety of stimuli depending on the authors' research questions. Earlier studies, such as Eilers, Wilson, and Moore's 1979 study, included up to 10 different consonant-vowel contrasts. Hillenbrand (1984) utilized consonant vowel contrasts such as /ma-na/, /ma-mi-mu/, and /na-ni-nu/ with both male and female speakers. Another study conducted by Hillenbrand and colleagues in 1979 utilizes the stimuli /be/, /we/, and /ue/. Other studies utilizing consonant-vowel contrasts include Nozza, Miller and Rossman's 1991 study, utilizing a /ba-ga/ contrast, and Goodsitt, Morse, Ver Hueve, and Cowan's 1984 study, which uses /ba/, /du/, /ko/, and /ti/ as its stimuli. Uhler and colleagues' recent studies, conducted in 2011, 2015, and 2018 also include consonant vowel contrasts as well as vowel contrasts.

VRISD studies also included differences in voice onset time as stimuli (Eilers, Gavin, & Wilson, 1979), vowel-consonant-vowel contrasts utilizing VRASPAC (Martinez et al., 2008; Eisenberg et al., 2004). Other examples of VRISD studies using multisyllabic stimuli include Bull, Eilers, and Oller's 1984 and 1985 studies. A summary of utilized stimuli can be seen in table 5, which is broken down into linguistic, phonemic, and other stimuli by procedure. OPP studies used a blend of linguistic, phonemic, and other stimuli while VRISD studies used primarily phonemic stimuli.

Stimulus	OPP	VRISD	Total
Linguistic	2	0	2
Phonemic	2	14	16
Other	5	1	6
Total	9	15	24

Table 5: Breakdown of stimuli by procedure.

Stimuli categorized as “other” can be seen in table 6. One VRISD study (Eilers et al., 1979) utilized differences in VOT while OPP studies utilized noise (Werner et al., 1992; Walker et al., 2019) and tone bursts (Sinnott and Aslin, 1985; Dasika et al., 2009; Cone and Whittaker, 2013).

Other Stimuli	OPP	VRISD
Differences in VOT	0	1
Noise	2	0
Tone bursts	3	0

Table 6: Stimuli categorized as other, organized by procedure

A summary of stimuli categorized as other can be found in table 6. Moreover, phonemic and linguistic stimuli are categorized as monosyllabic or multisyllabic, which can be seen in tables 7 and 8.

Phonemic Stimuli	OPP	VRISD
Monosyllabic	2	11
Multisyllabic	0	3

Table 7: Breakdown of phonemic stimuli by syllable count

Linguistic Stimuli	OPP	VRISD
Monosyllabic	0	0
Multisyllabic	2	0

Table 8: Breakdown of linguistic stimuli by syllable count

Stimulus level:

Another area of analysis includes stimulus presentation level across procedures. Of the 26 selected studies, 22 listed information on presentation level while 4 studies (Walker et al., 2019; Eisenberg et al 2004; Hillenbrand, 1984; Eilers et al., 1979) did not list this information. 5 studies (3 OPP, 2 VRISD) utilized masking noise. 3 studies (Greico-Calub et al., 2012; Uhler et al., 2018; Uhler et al., 2015) utilized multiple presentation levels, 2 of which utilized VRISD and one of which utilized OPP. Uhler and colleagues (2015;2018) utilized presentation levels or 50, 60, and 70 dBA and Greico-Calub and colleagues (2012) utilized presentation levels randomly roved +/-4 dB from a base intensity of 60 dB to minimize availability of monaural cues (p. 6). Dasika and colleagues (2009) implemented different presentation levels based on the cochlear implant manufacturer, with a presentation level of 60 dB for Advanced Bionics and Med El users and a presentation level of 65 dB for Cochlear users. The authors justified this approach based on estimated comfort levels provided by each manufacturer.

For remaining OPP studies, one study utilized a presentation level of 65 dB SPL (Lalonde and Werner, 2019), one used a level of 60 dB SPL (Sinnot and Aslin, 1985), and one utilized a level of 50 dB SPL (Cone and Whittaker, 2013). Of the remaining VRISD studies, 2 (Bull et al., 1984; Bull et al., 1985) utilized a presentation level of 60 dBC, one (Uhler et al., 2011) utilized a presentation level of 60 dB SPL, one (Cone, 2015) utilized a presentation level of 70 dB SPL, one (Hillenbrand, 1979) utilized a presentation level of 50 dBA, one (Eilers et al., 1977) used a presentation level of 50 dB SPL, and one (Martinez et al., 2008) utilized a presentation level of 70 dBA. Overall, presentation levels for OPP studies ranged from 30 dB SPL to 65 dB SPL while presentation levels for VRISD studies ranged from 50 dB SPL to 70 dB SPL. Table 9 shows presentation levels used by procedure.

	Procedure	
Stimulus	VRISD	OPP
30-40 dB	0	1
41-50 dB	4	1
51-60 dB	4	5
61-70 dB	2	1
multiple levels	2	1

Table 9: Presentation levels by procedure

Average Number of Visits:

Another important factor of analysis in selected papers is the number of visits needed to obtain results. For studies using OPP, 6 out of 9 studies listed information on the number of visits needed to obtain results. Number of visits ranged from 2.1 to 7.6 sessions, with an average of 3.8 visits between studies. 13 out of 16 studies using VRISD listed information on the number of visits required to obtain results. In studies utilizing VRISD, average number of visits ranged from 2 to 9.1 visits, with an average of 4.2 required visits between studies. It is noteworthy that in both procedures, the number of visits required from the year 2010 onward decreases to 3.5 visits for OPP procedures, and 2.0 visits for VRISD procedures, likely reflecting refinement of the procedures. Earlier studies required a larger number of visits, such as Sinnott and Aslin's 1985 OPP study (average 7.6 visits) and Hillenbrand's 1984 VRISD study (average 7.5 visits). Figure 4 shows average OPP visits over time while Figure 5 shows average VRISD visits over time. It is noteworthy that there is more data on number of visits provided in VRISD (13/15) papers compared to OPP (5/9) papers.

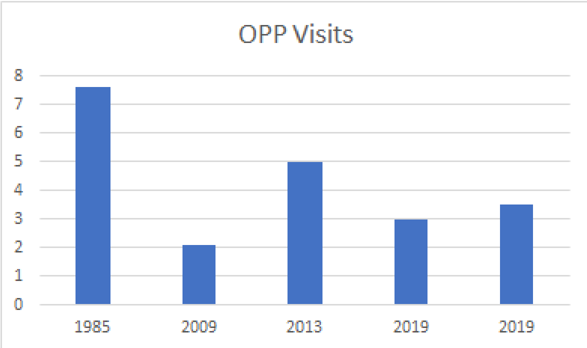


Figure 4: OPP visits over time

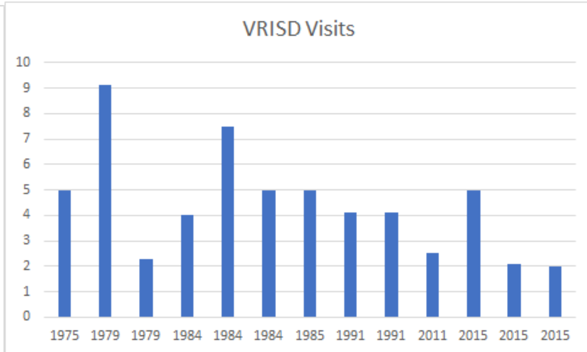


Figure 5: VRISD visits over time

Sample sizes:

Total sample sizes from studies ranged from 3 to 88, with an average sample size across studies of 26.5 infants. 5 OPP studies and 2 VRISD studies utilized adults as a control group. Leibold and colleagues (2016) also included 10 school-aged students in their study. VRISD samples ranged from 9 to 43 with an average sample size of 19.26 infants while OPP samples ranged from 7 to 88 with an average sample size of 39.3 infants. A two-tailed t-test of samples with unequal variances found significant differences between sample sizes and procedure ($p=0.028$). Figures 7 and 8 show sample sizes by procedure. Figures 7 and 8 show sample sizes by procedure.

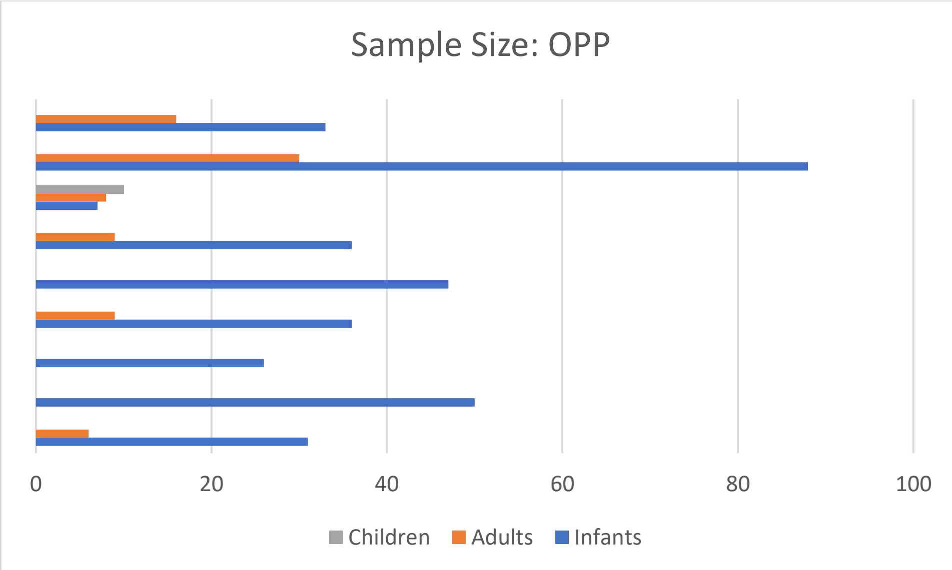


Figure 6: OPP sample sizes by research paper. Blue bars represent infant sample sizes, orange and gray bars represent adult and children sample sizes, respectively, if applicable.

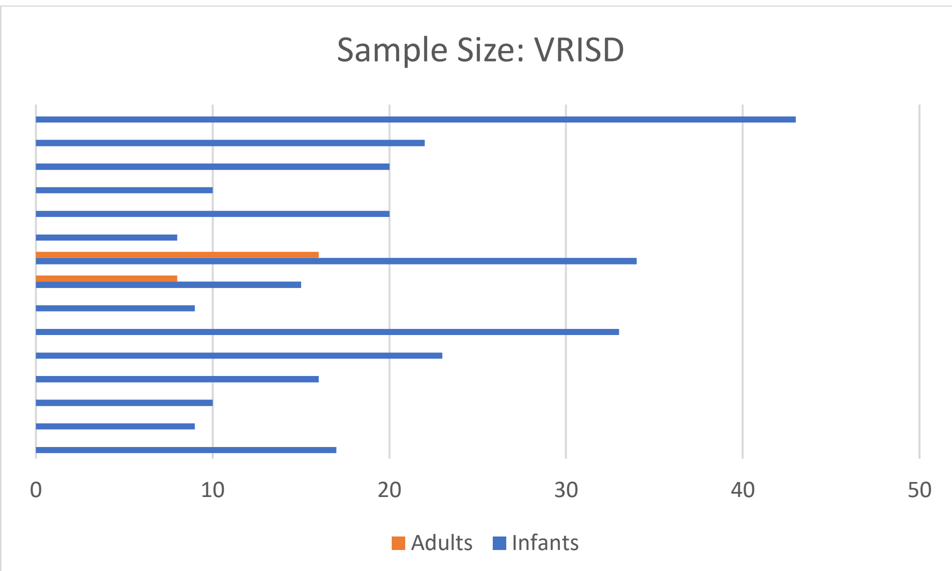


Figure 7: VRISD sample sizes by research paper. Blue bars represent infant sample sizes and orange bars represent adult sample sizes if applicable.

Shaping and training: OPP

When discussing the conditioning process in OPP testing, it is important to note that conditioning can be described in terms of the observer or the listener. It is also pertinent to account for differences in shaping and training procedures by researchers and desired outcomes in studies. Of 9 selected papers utilizing OPP, two (Greico-Calub et al., 2008; Greico-Calub et al., 2012) did not list shaping and training strategies. Sinnott and Aslin (1985) assessed frequency and intensity discrimination ability in infants and adults. Subjects were retrained at each session and shaping trials consisted of +/- 250 Hz stimuli when conditioning for frequency discrimination and +/- 15 dB when assessing intensity discrimination (Sinnott and Aslin, 1985). The reinforcer was activated for each shaping trial and was repeated until the infant anticipated the reinforcer. Two anticipatory head turns with no false alarms on the part of the observer were required to consider subjects trained (Sinnott and Aslin, 1985).

Werner and colleagues (1992) assessed gap detection ability in infants. Training was conducted utilizing gaps fixed at 100 ms. If the observer judged a signal trial had occurred when a signal trial did occur, the reinforcer was activated for 4 milliseconds. If the observer judged no signal when a signal trial had occurred, the reinforcer was activated, and an error was scored. Correct rejections and false alarms were also scored. The observer was considered trained when they correctly judged 4 out of 5 signal trials and 4 out of 5 no signal trials (Werner er al., 1992, p. 263). Dasika and colleagues (2009) utilized training and criterion trials phases prior to conducting testing. During the training phase, the reinforcer was activated on every signal trial, regardless of whether the observer correctly judged that a signal trial had taken place. Training lasted until 4 out of 5 signal trials were correctly observed with at least one correct rejection. The training phase lasted between 5 and 10 trials (Dasika et al., p. 8). During the criterion phase, there was a 0.5 probability of a signal trial. The reinforcer was only activated if the observer correctly identified a signal trial. This phase lasted until the observer correctly scored 4 out of 5 hits and 4 out of 5 correct rejections (Dasika et al., 2009, p. 8).

Cone and Whittaker (2013) utilized OPP with cortical evoked auditory potentials (CAEP) on a series tone bursts and speech tokens. For the OPP task, subjects were trained with up to 5 pairings of the target stimulus with the reinforcer. After 5 pairings, a probe signal trial was presented, with the reinforcer withheld until the infant responded within 4 seconds. The observer needed to complete 2 correct judgements of probe trials to proceed to testing (Cone and Whittaker, 2013, p. 12).

Leibold and colleagues' 2016 study consists of a conditioning phase and a training phase prior to the initiation of the test phase. In the conditioning phase and training phases, the target word is presented at "a level expected to be clearly audible" (Leibold et al., 2016, p. 348). The purpose of the conditioning phase is to familiarize listeners with the relationship between the target word and the reinforcer. There is a 0.8 probability of the trial presented to be a signal trial during this phase. The reinforcer is activated after each signal trial, regardless of whether the observer correctly judged that a signal trial had taken place. The conditioning phase is considered complete when the observer correctly responses to 4 out of 5 trials, including at least one correct rejection of a catch trial (Leibold et al., 2016, p. 348). The purpose of the training phase in this study is to demonstrate that the listener and observer could reliably complete the task. In this phase, the probability of a signal trial is 0.5. The training is considered complete after a run of 10 trials with a hit rate of 0.8 or higher and a correct rejection rate of 0.2 or lower. The authors note that with speech shaped noise, the average amount of needed trials to complete training is 20.0 for infants, 11.0 for children, and 11.9 for adults. Conversely, the average amount of trials to complete training utilizing a two-talker speech babble is 29.2 trials for infants, 10.2 trials for children, and 10.0 trials for adults. A large amount of variability in required training trials is noted for the infant group (Leibold et al., 2016, p. 348).

Lalonde and Werner's 2019 study utilizes familiarization and training phases prior to testing. The goal of the familiarization phase is to demonstrate the association between the signal and the reinforcer. The familiarization phase consists of 4 signal trials and one no-signal trial, with a speech stimulus presentation level of 14 dB SNR. Infants are reinforced for every trial regardless of their response to the stimuli (Lalonde and Werner, 2019, p. 3864). The goal of the training phase in this study is to establish that the observer can reliably identify that a signal trial has occurred. In the training phase, the speech stimuli are presented at 10 dB SNR, and signal and no-signal trials are presented in random order. The training phase continues until the observer correctly identifies 4 out of 5 signal trials and 4 out of 5 no-signal trials. If participants did not reach criterion within 40 trials, the listener was given a break or testing was continued another day. Lalonde and Werner (2019) note that the average amount of trials required to complete training are 19.9 for the study's detection task and 19.5 for the study's discrimination task (p. 3864).

Walker and colleagues' 2019 study consists of training and criterion phases prior to the initiation of the testing phase. The authors assess modulation depth (change in loudness) and modulation rate (change in frequency) in infants. The utilized stimuli in both training and criterion phases are a stimulus with 0 dB modulation depth and and +/- 10 Hz modulation rate (Walker et al., 2019, p. 3670). In the training phase, there is a 75% chance of a signal trial occurring, and the reinforcer is activated despite the listener's response. To complete the training phase, the observer must obtain a hit rate of 80% or greater and a false alarm rate of 80% or less for the last 5 signal and no signal trials. If this was not completed within 20 trials, the listener was retrained. The criterion phase was utilized to confirm that the listener and observer could detect modulated stimuli at "maximal modulation depth" (Walker et al., 2019, p. 3670). Criterion training was completed in 20 trial blocks, with a 50% chance of a signal trial occurring. The reinforcer was activated only when the observer correctly judged that a signal trial had occurred.

If the criterion phase was not completed in 40 trials, it was repeated once. If this phase could not be completed after the second attempt, no further testing was completed and no threshold was obtained for that subject (Walker et al., 2019, p. 3670).

Overall, 5 of 7 OPP studies that listed training procedures required the observer to correctly judge the presence of a target trial at least 4 out of 5 consecutive times prior to beginning testing. The remaining studies noted that two anticipatory head turns to a change in stimulus are required before initiation of testing (Sinnott & Aslin, 1985) and two correct judgements of a change trial (Cone & Whittaker, 2013).

Shaping and training: VRISD

Of the 16 selected articles utilizing VRISD, 12 listed shaping and training procedures in detail while 4 (Eisenberg et al., 2004; Uhler et al., 2011; Uhler et al., 2015; Uhler et al., 2018) did not. It is notable, however, that while Uhler and colleagues' 2015 and 2018 studies did not conduct training trials, if the infant did not reach criterion in the initial condition of 50 dBA, they were given additional conditions of 60 and 70 dBA. This protocol may have potential for clinical practice as it can be more time efficient and allows for additional testing only when warranted.

Eilers and colleagues (1977) utilized training trials by presenting the target stimulus at 65 dB SPL with the background stimulus presented at 50 dB SPL. If the infant turned their head towards the speaker, a visual reinforcer was delivered for two seconds. If the infant did not initiate a head turn after 2 trials, the assistant in the booth pointed to the reinforcer and verbally encouraged the infant to look towards the reinforcer. The authors note that after 2-3 trials of training, most infants would turn their heads towards a change in stimulus (Eilers et al., 1977, p. 773). A 45-degree head turn was required for a response to be scored and eye movement or change in arousal was not scored as a response. The level difference between the target and background stimulus was gradually reduced over 5 dB steps until the target and background stimulus were each 50 dB SPL (Eilers et al., 1977, p. 773). Hillenbrand and colleagues (1979) paired the reinforcer with a change in stimulus to elicit a head turn response in training. During this process, the target stimulus was presented at a level of 70 dBA and the background stimulus was presented at a level of 50 dBA. After each correct head turn, the stimulus was decreased by 5 dB until the target and background stimulus reached the same level of 50 dBA. A total of 3-29 trials were required for subjects during training, with an average of 9.1 trials between subjects (Hillenbrand, 1979, p. 153). Eilers and colleagues (1979) referenced Eilers and colleagues' 1977 study when describing training procedures. However, this study utilized differences in voice onset time as stimuli; therefore, infants were trained with the most natural sounding stimulus (bit vs. beat) as their training stimulus.

Goodsitt and colleagues (1984) describe a training phase which occurred in the first session of testing while the remaining trials were conducted in the 2 to 3 subsequent testing sessions. The training stimulus pair, /ba-du/, was counterbalanced between subjects as to which sound was the background stimulus and which sound was the target stimulus. All trials were change trials during the training phase. The target was presented 12 dB louder than the background stimulus and after 2 consecutive head turn responses, the level was gradually decreased by 4 dB until both the target and background stimuli reached the testing level of 54 dB. Following training, a criterion phase was initiated in which the probability of a change trial occurring was 0.5. For the reinforcer to be activated, both the experimenter and the assistant located in the booth needed to agree that a head turn response was elicited on change trials. 9 out of 10 correct responses were required by the infant in both change and control trials to be included in the experiment (Goodsitt et al., 1984, p. 904-905).

Hillenbrand (1984) notes that a head turn response was conditioned by initiating a change trial and activating the reinforcer. For subjects to be included in the study, they needed 3 consecutive anticipatory head turns to a change in stimulus and were allowed a maximum of 25 trials to meet conditioning criteria. Additionally, Hillenbrand (1984) notes that re-training occurred when performance dropped. This included when infants missed change trials 3 consecutive times, or if at the end of 15 trials the infant missed half of the change trials, 5 additional training trials were conducted (Hillenbrand, 1984, p. 1616). Bull, Eilers, and Oller (1984) utilized a protocol in which each session began with a shaping phase with the contrast stimulus presented 12 dB louder than the background stimulus. The reinforcer was activated when the infant produced a head turn in response to a change in stimulus, however if no response was elicited within 6 seconds of presentation, the reinforcer was elicited regardless of response. Once the infant responded correctly on 2 consecutive trials, the level of the contrast stimulus was gradually increased by 4 dB until both the target and background stimuli were presented at the testing level of 60 dB (Bull et al., 1984, p. 15). Bull, Eilers, and Oller's 1985 study has an identical training protocol to their 1985 study.

Nozza, Miller, Rossman, and Bond (1991a) describe shaping and training phases prior to the initiation of testing. In the shaping phase, a repeating background stimulus, /ba/, is presented at 70 dB SPL in quiet. All trials in this phase were change trials, with /ga/ serving as the stimulus contrast presented at 80 dB. Once an anticipatory head turn was produced for 2 consecutive trials, the level of the contrast stimulus was decreased in 5 dB steps with the same criterion for a decrease in level until both the background stimulus and contrast were 70 dB. Once the stimuli were at the same level, 3 consecutive correct head turns were needed to continue to the training phase. In the training phase, masking noise was introduced with a 0.5 probability of a trial being a change trial. In order to proceed to testing, the infant must have either 8 out of 10 correct responses or 14 out of 20 correct responses (Nozza et al., 1991a, p. 646). Nozza, Rossman, and Bond's 1991(b) study follows an identical shaping and training protocol, however masking is introduced during the testing phase.

Martinez and colleagues (2008) note that a string of vowel-consonant-vowel utterances, with the standard being "oodoo", were presented at 70 dBA until the infant habituated to the standard stimulus. When the stimulus changed to the target stimulus, the infant was conditioned to look towards the reinforcer. 2-3 trials were generally required to complete conditioning (Martinez et al., 2008, p. 2-3). Cone (2015) compares electrophysiological results obtained via mismatch negativity/P300 to behavioral speech discrimination results in a group of infants. The author describes training as 5 parings of change trials with the reinforcer followed by probe trials with a stimulus change. During probe trials, the reinforcer was withheld until the child responded. The infant needed to correctly respond to 2 out of 3 probe trials to move on to testing. If the infant was not conditioned after probe trials, 5 more training trials followed by 2 more probe trials were conducted. Up to 15 training trials were permitted per session, and the infant was excluded from this portion of testing if they did not meet criterion (Cone, 2015, p. 8). Interestingly, response criteria included eye widening or looking towards the reinforcer, or a head turn, which is not characteristic to VRISD.

Of the 12 studies that listed shaping and training procedures, 8 describe their shaping and training procedure involving the presentation of the target stimulus at a higher sensation level compared to the background stimulus. The target stimulus is then lowered in decrements until it is the same level as the background stimulus. This is not unlike training for VRA, when the clinician will play a stimulus at an audible level to condition the child for threshold testing,

Risk of Bias:

Of the 9 selected papers utilizing OPP, 4 papers did not list information on potential for bias in results (Sinnott & Aslin, 1985; Dasika et al., 2008; Cone & Whittaker, 2013; Walker et al., 2019). Werner and colleagues (1992) found that all infants had a higher false alarm rate versus adults, suggestive of a response bias in infants (p. 266-267). An Arcsin transformation was performed to eliminate the effects of response bias, however the authors noted that differences in response bias cannot account for differences between results obtained 500 Hz versus other masking conditions in infants (Werner et al., 1992, p. 266-267). Greico-Calub and colleagues' 2008 and 2012 studies utilized masking music for assistants and caregivers in the audiometric booth to eliminate inadvertent cueing to the infant. Leibold and colleagues (2016) found age effects in response bias, with adults and school-aged children having a more conservative response bias compared to infants (p. 350). Authors noted that it is possible that differences in response bias between groups may be responsible for the pattern of threshold differences between masker conditions for each group (Leibold et al., 2016, p. 352). Lalonde and Werner (2019) report that masking was utilized on adults in the audiometric test booth to prevent inadvertent cueing to the infant.

Of 15 selected VRISD papers, 6 did not include information on potential biases (Eilers et al., 1977; Eilers et al., 1979; Goodsitt et al, 1984; Martinez et al, 2008; Uhler et al., 2011; Cone, 2015). Hillenbrand and colleagues (1977; 1979), Bull and colleagues (1984; 1985), Eisenberg and colleagues (2004), Uhler and colleagues (2015), and Uhler and colleagues (2018) mention utilizing masking on adults located in the test booth via headphones to prevent inadvertent cueing of the infant during testing. Nozza and colleagues (1991a) counterbalanced stimuli to decrease the chance of bias; however, the authors noted that since no subject was screened for hearing loss prior to testing the inadvertent inclusion of infants with hearing loss cannot be ruled out. Nozza and colleagues (1991b) discuss the possibility of selection bias in their studies.

Testing time:

An important aspect of implementing a behavioral speech discrimination procedure clinically in infants is required time for testing. Six VRISD and two OPP papers list some amount of information on amount of stimulus presentations or number of trials in a testing session, which are listed in Table 9. Overall, more data on testing time is present in VRISD papers. Of VRISD papers that list number of trials utilized per speech sound contrast (n=5), the average number of trials presented per contrast is 20.4. From Walker and authors (2019) and Lalonde and Werner (2019), recent OPP studies have utilized sessions of 20 trials. Data on testing time from earlier OPP procedures is not listed.

Study	Procedure	Information on testing time:
Hillenbrand, 1984	VRISD	20 trials per session, sessions last 10-15 minutes, 7-8 sessions required to complete testing
Bull et al., 1984;1985	VRISD	30 trials per contrast
Uhler et al., 2015;2018	VRISD	15 trials per contrast
Cone, 2015	VRISD	12 trials per contrast
Walker et al., 2019	OPP	Testing done in blocks of 20 trials per condition; 2 conditions for testing
Lalonde and Werner, 2019	OPP	10 signal and 10 no signal trials per session

Table 10: information on length of testing by study and procedure

Results obtained: OPP

Sinnott and Aslin (1985) found that when utilizing frequency increments while testing infants, 18 out of 38 subjects produced valid difference limens while with the use of frequency decrements, 27 out of 37 infants produced valid difference limens. Moreover, infant difference limens ranged between 11 and 29 Hz while difference limens for adults ranged from 3 to 5 Hz. No significant differences between increment and decrement conditions were noted. With intensity increments, 26 out of 56 sessions produced valid difference limens while with intensity decrements, 0 out of 21 sessions produced valid difference limens. Infant difference limens regarding intensity ranged from 3 to 12 dB, while adult difference limens ranged from 1.5 to 2.0 dB (Sinnot and Aslin, 1985, p. 1990-1991). This provides evidence that infants require a wider difference in frequency stimuli to recognize a difference in frequency and that infants also require a larger amount of intensity to recognize intensity differences.

Werner and colleagues (1992) found that infants perform worse than adults with gap detection tasks utilizing OPP. 3- and 6-month-old infants have gap detection thresholds that are 40 to 60 milliseconds higher than that of adults in all masking conditions. Conversely, the effect of frequency on gap detection threshold is similar for 3- and 6-month-olds and adults in that threshold improves as frequencies are made available up to 8kHz. Performance of 12-month-old infants differs from both older and younger listeners. 12-month-olds performed no better than 3- and 6-month-olds with 2kHz and 8kHz masker cutoff conditions and when the masker cut-off frequency was as low as 500 Hz, 12-month-olds obtained gap detection thresholds like that of adult listeners (Werner et al., 1992, p. 264-265). This shows that 12-month-olds exhibit some degree of auditory maturation in a gap-detection task, however it varies by masker cutoff frequency.

In Greico-Calub and colleagues' 2008 study, only half of bilateral cochlear implant user could produce minimal audible angles (MAA) in localization tasks while no unilaterally implanted child could perform a right-left discrimination task above chance (p. 239) Additionally, the authors found that when compared to normal hearing peers, children with bilateral cochlear implants show large inter-subject variability in their performances. Moreover, 3 toddlers with bilateral cochlear implants seem to have age-appropriate localization acuity in fixed level and intensity roved conditions. Differences between unilaterally and bilaterally implanted subjects' performances cannot be attributed to differences in chronological age or differences in hearing experience (time from initial activation) considering that the unilateral group was at a slight advantage for each of these categories (Greico-Calub et al., 2008, p. 239). Results from this study provide evidence that bilateral input through cochlear implants can facilitate the development of spatial hearing at a young age.

Dasika and colleagues (2009) found that 11 out of 12 infants utilizing cochlear implants (CI) were able to complete the task, all of whom reached criterion after the first visit with a median of 14 trials. 10 out of 11 infants that completed the task required 26 or fewer trials. Interestingly, children with cochlear implants reached criterion quicker than normal hearing infants. With CI users, there was no correlation between age and amount of experience in results obtained. Children's detection ability improved with increased level and those with CI exhibited steeper psychometric function slopes than normal hearing listeners. Additionally, reaction time for false alarms on no signal trials were significantly greater than reaction times for hits on signal trials when tested at the highest intensity level (Dasika et al., 2009, p. 14). This suggests that longer reaction times in OPP testing may be suggestive of no-signal trials that the observer incorrectly judges as a signal.

Greico-Calub and colleagues (2012) found that infants with normal hearing had significantly better localization acuity compared to infants with unilateral CI and bilateral CI. Children with bilateral CI's performed significantly better in localization tasks, than with their first CI alone and children with over 12 months of CI experience performed better than those with less than 12 months experience. This contrasts with Dasika and colleagues' 2009 study, which found no correlation with CI experience and performance on their tasks. However, Dasika and colleagues studied reaction time while Greico-Calub and colleagues studied localization ability. Results from these two studies can allow us to infer that amount of CI experience can have different effects on different tasks.

Cone and Whittaker (2013) observed that threshold differences were significantly greater in infants compared to adults. Thresholds for speech sounds in infants were obtained at an average of 36 dBA SPL while thresholds for adults were obtained at 10 dbA SPL. Infant thresholds to speech tokens did not vary with speech sound, consonant versus vowel stimulus, or manner of consonant production while adult thresholds for /u/, /m/, and /s/ tokens were elevated compared to /i/, /a/, /o/, and /u/ speech tokens. Additionally, Cone and Whittaker (2013) found that in infants, thresholds for speech tokens were 10-13 dB higher than thresholds for tone bursts and that CAEPs were robust at levels like that of behavioral thresholds (p. 16).

Leibold and colleagues (2016) found that the average adult threshold was approximately 24 dB lower on average than infants in both masking conditions (p. 348). Analysis of variance group found a significant effect for masker type, age group, and masker type x age group (Leibold et al., 2016, p. 349). Additionally, the authors found that 8-10-month-old infants had substantially more difficulty detecting disyllabic words in the presence of speech-shaped noise, or a 2-talker babble compared to 8-10-year-old children and adults (Leibold et al., 2016, p. 350).

Lalonde and Werner (2019) found a significant difference between results of auditory-only and auditory-visual conditions (p. 3866). The authors note that infants benefit from auditory-visual cues in both detection and discrimination tasks. Results from this study suggest that 6-8.5-month-old infants are relatively mature in their ability to use visual onset-offset cues to detect speech in noise but are still developing the ability to utilize phonetic cues (Lalonde and Werner, 2019, p. 3868). Walker and colleagues (2019) found that infants required 10 dB greater modulation depth to detect a 10 Hz modulation compared to adults, which was significant via a two-tailed independent samples t-test (p. 3671). Moreover, infants were less sensitive than adults at detecting amplitude modulation, but temporal modulation cutoff frequencies did not significantly differ from adults (Walker et al., 2019, p. 3671-3672). Table 11 displays the age range and results obtained relating to speech discrimination ability in OPP studies. Studies that are bolded utilized infants with hearing loss in the subject pool.

Study	Procedure	Age Range	Ability
Werner et al., 1992	OPP	3 months	infants demonstrate gap detection ability
Werner et al., 1992	OPP	6 months	infants demonstrate gap detection ability

Lalonde et al., 2019	OPP	6-8.5 months	Infants relatively mature in ability to use onset/offset visual cues in a speech discrimination task
Sinnott et al., 1985	OPP	7-9 months	Infants demonstrate frequency and intensity discrimination
Leibold et al., 2016	OPP	8-10 months	Infants can detect bisyllabic and trisyllabic words in two-talker babble and speech-shaped noise
Werner et al., 1992	OPP	12 months	Gap detection ability similar to 3- and 6-month-olds with higher masker cutoff frequencies, and similar to adults in conditions with lower masker cutoff frequencies
Greico-Calub et al., 2008	OPP	**26-36 months**	**50% of bilateral cochlear implant users can produce minimal audible angles compared to 0% of unilateral users**

Greico-Calub et al., 2012	OPP	**26-36 months**	**Children with over 12 months CI experience perform better on a localization task**

Table 11: Results by study and age group in OPP papers. Bold text indicates that the study utilized children with hearing loss in the subject pool.

Results obtained: VRISD

Eilers and colleagues (1977) tested a variety of consonant-vowel and vowel-consonant contrasts in normal hearing infants aged 1-3 months old, 6-8 months old, and infants aged 12 months. The authors found that all age groups could distinguish between /va/ and /sa/, /sa and /sha/, /as/ and /a:z/, and /at/ and /a:d/ while no group could successfully distinguish /fa/ from /tha/, /a:t/ from /at/, or /fi/ from /thi/. Infants aged 12 months could successfully distinguish /at/ from /a:d/ and /at/ from /a:t/, while only one infant in this age group could discriminate /fa/ from /tha/ (Eilers et al., 1977, p. 20-21). The authors do note variability among subjects, and therefore this data should be interpreted with caution. Hillenbrand (1979) utilized /be-we/, /be-ue/, and /we-ue/ contrasts in two separate groups, both aged 6-7 months. One group was given recorded speech stimuli while the second group received synthetic speech stimuli. In the first group (synthetic stimuli), /be-we/ and /be-ue/ were successfully discriminated in trials, while /we-ue/ was not while in the second group (recorded stimuli), discrimination performance for all contrasts was significantly above chance (Hillenbrand, 1979, p. 155, 159).

Eilers and colleagues (1979) assessed voice onset time (VOT) discrimination ability in English and Spanish language learning infants. A 2x2 ANOVA with repeated measures was performed with the between subjects variable being language experience and the within subjects variable being the stimulus pair which yielded no significant overall effect for language experience (Eilers et al., 1979, p. 16). A significant main effect for language experience was obtained ($p < 0.01$). Additionally, language experience by stimulus pair interaction was also significant ($p < 0.01$). In Spanish learning infants, participants had an average of 80% correct head turns for the lead pair and 86% correct for the lag pair while English learning infants had an average of 46% correct head turns for the lead pair and 92% correct for the lag pair (Eilers et al., 1979, p. 16). This data loosely suggests that speech sound discrimination abilities can differ based on language experience. Goodsitt and colleagues (1984) utilized consonant vowel speech signals /ba/, /du/, /ko/, and /ti/, with /ba/ and /du/ being the target syllable in redundant (i.e.: /tibati/) and mixed (i.e.: /tibako/) trisyllabic conditions. A 2x2 ANOVA revealed significant effects for context structure (redundant vs. mixed) and target syllable (ba vs. du). Redundant contexts, meaning that the target syllable is presented with two of the same background syllables, yielded significantly higher results ($p <0.005$) and the target syllable /ba/ yielded significantly higher results ($p<0.005$) compared to /du/ as a target syllable (Goodsitt et al., 1984, p. 906-907).

Hillenbrand (1984) found significant main effects for trial type (change vs. control) and group (phonetic vs. non-phonetic) with a significant interaction between group and trial type (p. 1616-1617). The author noted that “infants perform well only when stimuli can be organized along some salient dimension” and that some infants responded less to stimuli with /i/ included (Hillenbrand, 1984, p. 1616-1617). However, there is a large amount of variability in results, therefore the data should be interpreted with caution. Bull and colleagues (1984) found that discrimination scores were higher overall for two-syllable stimuli compared to 3-syllable stimuli and that discrimination from +2dB final syllable increments were significantly poorer than +4 or +6 dB final syllable presentations (p. 15-16). Bull and colleagues’ 1985 study found that adults had overall higher discrimination ability compared to infants in both bisyllabic and trisyllabic conditions (p. 293).

Nozza, Rossman, and Bond (1991a) obtained a /ba-da/ discrimination threshold average of 39.3 dB SPL for infants and 11.8 dB SPL for adults and an average /ba-ga/ discrimination threshold of 35.1 dB SPL in infants and 9.7 dB SPL for adults (p. 107). Additionally, main effects for age group and contrast were noted. Nozza, Miller, Rossman, and Bond (1991b) assessed validity and reliability of VRISD utilizing a /ba-ga/ contrast in noise. In this study, infants had an average discrimination threshold 46.3 dB SPL (SD = 4.5) and adults had an average discrimination threshold of 39.4 dB SPL (SD = 1.4) (Nozza et al., 1991b, p. 647). The authors note that infants require a signal-to-noise ratio of –1.7 dB on average while adults require a signal-to-noise ratio of –8.6 dB on average to discriminate /ga/ from /ba/ in noise. Utilizing a t-test for differences between groups, the differences in results between infants and adults are statistically significant (Nozza et al., 1991b, p. 647). In assessing test-retest reliability with VRISD, the authors note that the average difference between two testing sessions in infants is 1.8 dB (SD = 6.7) while differences between two testing sessions in adults was 0.3 dB (SD = 3.3) on average in adults. Differences between groups concerning test-retest reliability were not significant. Additionally, the authors noted that 14/16 (87.5%) of infants tested had a test-retest threshold difference of 10 dB or less while adults had test-retest differences of 5 dB or less (Nozza et al., 1991b, p. 647). Results from this study align with previous studies, suggesting some degree of external validity in these results.

Eisenberg and colleagues (2004) provide pilot data utilizing the Visual Reinforcement Assessment of the Perception of Speech Pattern Contrasts (VRASPAC) procedure. The authors tested 5 children with normal hearing, 2 children with cochlear implants, and 1 child with mild hearing loss utilizing hearing aids. Children with normal hearing responded consistently to all contrasts while those with hearing loss were able to consistently respond to vowel contrasts only (Eisenberg et al., 2004, p. 365). The authors note limitations to this data, including small sample size, no masking of parent/caregiver or tester, possible cueing from the tester, and false positives that were not taken into consideration in data analysis. Considering this, this preliminary data should be interpreted conservatively.

Martinez and colleagues (2008) utilized VRASPAC to assess speech contrasts in infants aged 7-21 months with normal hearing and with varying degrees of hearing loss ranging from mild to profound. In infants with normal hearing, vowel height (oo-ah) and vowel place (oo-ee) contrasts were discriminated with high levels of confidence (>90%). All but one infant with hearing loss were able to discriminate vowel height contrasts with a high level of confidence (Martinez et al., 2008, p. 3-4). With the vowel height contrast, infants with hearing loss less than 60 dB HL could discriminate with a high level of confidence while those with hearing loss greater than 60 dB HL could not. The method of quantifying hearing loss is not specified when describing hearing loss greater or less than 60 dB HL. Additionally, the authors found that consonant contrasts had variable discrimination scores regardless of hearing level or contrast (Martinez et al., 2008, p. 3-4).

Uhler and colleagues (2011) conducted a longitudinal study on 3 children who received cochlear implants utilizing VRASPAC with 7 normal hearing children included as a control group. The authors found that inter-rater reliability was good overall (Uhler et al., 2011, p. 136). It was found that newly implanted toddlers could discriminate 3 out of 5 phonemic contrasts 3 months after the initial stimulation, suggesting that 13 weeks is a long enough time post-initial stimulation for children to discriminate the utilized speech sound contrasts (Uhler et al., 2011, p.

139). Compared to normal hearing peers, 2 out of 3 cochlear implant recipients performed no differently on at least 3 out of 5 speech sound contrasts (Uhler et al., 2011, p. 139). Unsurprisingly, the child who had the most residual hearing prior to implantation had the best performance of the three subjects. While these results are promising for speech discrimination ability in children utilizing cochlear implants, the sample size is small, and the results therefore cannot be generalized to a specific population.

Cone (2015) compared VRISD behavioral speech discrimination results with cortical auditory evoked potential (CAEP) results in 20 normal hearing infants. Overall, a hit rate of 68.5% and a false alarm rate of 25.7% was obtained. It is noted that for 28.1% of trials, the false alarm rate was 50% or greater (Cone, 2015, p. 12). The contrast with the highest average hit rate (77.1%) was /a-u/ while the contrast with the lowest average hit rate (58.8%) was /a-o/. CAEP results from /a-u/ had a correlation of 0.76 with behavioral responses pertaining to P1-N1 amplitude, suggesting that this contrast has "more perceptual salience" compared to others (Cone, 2015, p. 12-13). Waveforms obtained from /a-u/ contrasts were larger and had shorter latencies compared to other contrasts used. However, the author mentions that this data is limited, and that results should be interpreted with caution.

Uhler and colleagues (2015) assessed 22 normal hearing infants' discrimination of /a-i/ and /ba-da/ contrasts utilizing VRISD. Infants were initially assessed at 50 dbA and if the infant reached a criterion of 75% correct or more, they did not receive testing conditions at 60 and 70 dbA. The authors found that 62% of infants reached criterion at 50 dBA for the /a-i/ contrast, with an additional 14% and 19% of infants reaching criterion at 60 and 70 dBA respectively (Uhler et al., 2015, p. 6). One infant (4.5%) did not reach criterion at any presentation level. For the /ba-da/ contrast, 38% of infants reached criterion at 50 dbA, with an additional 14% and 19%

reaching criterion at 60 and 70 dbA respectively (Uhler et al., 2015, p. 6). 6 infants (27.2%) did not reach criterion at any presentation level for the /ba-da/ contrast. Overall, 95% of infants reached criterion for the /a-i/ contrast with an average intensity of 50.83 dB SL while 71% of infants reached criterion for the /ba-da/ contrast with an average intensity of of 56.36 db SL (Uhler et al., 2015, p. 6-7). This data suggests that normal hearing infants need a higher presentation level to discriminate the consonant contrast /ba-da/.

Uhler and colleagues (2018) replicated their 2015 study utilizing /a-i/ and /ba-da/ contrasts in children with hearing loss. With /a-i/, 50% of infants with hearing loss reached criterion at 50 dBa, with 95% of infants reaching criterion at any presentation level utilizing amplification (Uhler et al., 2018, p. 6). No significant differences between infants with hearing loss and infants with normal hearing evaluated in the 2015 study with the /a-i/ contrast. This suggests that presentation level's effect on speech discrimination for a vowel contrast such as /a-i/ is similar for both groups. With the /ba-da/ contrast, only 50% of infants with hearing loss could reach criterion at any level while aided. While the proportion of infants reaching criterion for /ba-da/ at 50 dBA did not significantly differ between normal hearing and hearing loss groups, children with normal hearing were 21% more likely to reach criterion for /ba-da/ at any level (Uhler et al., 2018, p. 6). The proportion of infants with hearing loss reaching criterion was significantly higher for the /a-i/ contrast compared to the /ba-da/ contrast ($p = 0.0004$) while this proportion was not significant for normal hearing infants ($p = 0.45$) (Uhler et al., 2018, p. 6).

Table 12 displays the same data for VRISD studies. Bolded text indicates that children with hearing loss were utilized in the subject pool.

Study	Procedure	Age range	Ability
Cone, 2015	VRISD	4-11.8 months	Infants able to discriminate /a-o/, /a-u/, and /a-i/ contrasts
Bull et al., 1984;1985	VRISD	5-11 months	Better discrimination scores in bisyllabic compared to trisyllabic contexts
Hillenbrand, 1984	VRISD	5.5-6.5 months	Infants able to discriminate /m/ from /n/ in various CV contexts
Eilers et al., 1977	VRISD	6 months	Infants able to discriminate /va-sa/, /sa-sha/, /as-a:z/, and /at-a:d/
Eilers et al., 1979	VRISD	6 months	Effect of language experience on voice-onset time perception observed

Goodsitt et al., 1984	VRISD	6.5 months	Performance above chance in discriminating a CV target within a trisyllabic string of syllables. Better performance with /ba/ as target compared to /du/. Better performance in redundant contexts
Uhler et al., 2015	VRISD	6-14 months	21/22 infants able to discriminate /a-i/, 16/22 infants able to discriminate /ba-da/
Nozza et al., 1991b	VRISD	7-11 months	Infants able to discriminate /ba-ga/ contrast in noise
Martinez et al., 2008	VRASPAC	7-17 months	Infants able discriminate vowel height and vowel place contrasts; inconsistent consonant discrimination results
Uhler et al., 2011	VRASPAC	7-18 months	Infants able to discriminate /a-u/, /a-i/. /u-i/, /ta-da/, and /pa-ka/ contrasts

Uhler et al., 2018	**VRISD**	**7-28 months**	**95% of infants with hearing loss able to discriminate /a-i/, 50% of infants with hearing loss able to discriminate /ba-da/**
Nozza et al., 1991a	VRISD	9-11 months	Infants able to discriminate /ba-ga/ contrast in quiet
Eisenberg et al., 2004	VRASPAC	9-34 months	Infants successfully discriminate /daa-doo/, /dee-doo/, /too-doo/, /zoo-doo/, /boo-doo/, and /goo-doo/
Martinez et al., 2008	**VRASPAC**	**9-21 months**	**Infants with hearing loss can discriminate vowel height contrasts and those with hearing loss better than 60 dB are more likely to discriminate vowel place; inconsistent consonant discrimination results**
Eisenberg et al., 2004	**VRASPAC**	**11-38 months**	**Infants with hearing loss can discriminate vowel height (/daa-doo/) contrasts**

Uhler et al., 2011	**VRASPAC**	**12-16 months**	**2 of 3 CI recipents performed similarly on 3 of 5 contrasts (included contrasts: /a-u/, /a-i/. /u-i/, /ta-da/, and /pa-ka/)**

Table 12: Results by study and age group in VRISD papers. Bolded text indicates that the study included children with hearing loss in the subject pool.

Table 13 summarizes quantitative results across studies including age, number of visits required to complete testing, and sample size. Two-tailed t-tests of samples with unequal variances were applied for each variable. A significant difference between sample sizes is observed, however there are no significant differences between age of subject and number of visits.

	OPP	VRISD	Significance
Age Range	3-36 months	5-34 months	--
Average Age	12.93 months	9.86 months	p=0.41
Visits range	2.1-7.6 visits	2-9.1 visits	--
Average Visits	4.24 visits	4.43 visits	p=0.86
Sample size range	7-88 subjects	9-43 subjects	--
Sample size average	39.3 subjects	19.26 subjects	**p=0.028**

Table 13: Comparison of age range, number of visits, and sample size by procedure with averages and t-test results

DISCUSSION

The primary goal of this systematic review is to evaluate VRISD versus OPP in terms of potential clinical utility using information from published studies meeting the inclusion criteria in terms of the procedures, results, and simplicity for an audiological setting. Results from infants with normal hearing versus infants with hearing loss were also examined. All studies included in this review utilized either OPP or VRISD as its primary procedure, included infants under the age of 36 months, and assessed discrimination ability.

Infant versus Adult results:

Adults have been included in some of the studies reviewed here as a control. Overall, results from VRISD and OPP studies have highlighted some trends. Of note, infants generally require a higher presentation level than adults to successfully discriminate between auditory stimuli. This is evidenced in the results of Nozza and colleagues (1991) and Bull and colleagues (1984; 1985), both of which utilized VRISD, where adults performed better than infants at the same task. Additionally, Sinnott and Aslin (1985) and Werner and colleagues (1992) utilized OPP, and results indicate that infants have poorer intensity and loudness discrimination ability and poorer gap detection ability compared to adults, respectively. This is not particularly surprising, as prior studies involving pure tone detection found that infants need higher sensation levels to respond to a sound behaviorally. For example, Tharpe and Asmead (2001) found that 3-month-old infants acquired behavioral thresholds 20 dB higher than adults and 6-month-old infants acquired behavioral thresholds around 10 to 15 dB higher than adults on average (p. 108). These results have implications for speech perception, since the ability to discriminate intensity and temporal information underlies speech perception.

Werner and Boike (2001) noted that infant pure tone thresholds at 1000 Hz are 15 dB poorer than adults in quiet while thresholds are 8 dB poorer than adults in noise (p. 2617). The authors cite Keefe and colleagues (1993), who reported that the difference between infant and adult pure tone thresholds in quiet can be partially attributed to middle ear immaturity. Moreover, results from Werner and Boike's study found that infants thresholds for detecting broadband noise are comparable to that of adults (Werner and Boike, 2001, p. 2109). This provides evidence against the idea that infants are simply inattentive during pure tone testing and suggests that infants develop better pure tone detection abilities within the first year of life.

Considering that infants require a higher presentation level to detect pure tones and that the auditory system continues to develop after birth, it is not unsurprising that infants have more difficulty discriminating between speech contrasts than adults. Of the reviewed studies, 6 OPP studies and 2 VRISD studies utilized adults as a control group. All studies that utilized adults as a control group found significant differences in results between infant and adults.

Differences between stimuli:

There are general differences in stimuli by procedure. In VRISD procedures, 10 out of 14 studies utilized consonant-vowel (CV) stimulus contrasts while only 1 of 9 OPP studies utilized a CV contrast, and results on discrimination of individual stimuli were not discussed in depth. In evaluating potential utility of VRISD, the consistent use of stimuli is beneficial as it allows for analysis of results across studies. Stimuli used in OPP papers were more varied, including frequency and intensity discrimination ability (Sinnot & Aslin, 1985), gap detection ability (Werner et al., 1991), spondee words used to assess localization ability (Greico Calub et al, 2008;2012), reaction time (Dasika et al., 2009), detection of spondee words in different masking conditions (Leibold et al., 2016), and modulation detection ability (Walker et al., 2019).

Considering stimuli used and dependent variables used in selected OPP papers, one can loosely infer that the procedure may be used clinically to assess abilities beyond speech discrimination ability. However, without consistent stimuli and dependent variables across procedures, it is difficult to assess combined results from OPP papers.

It appears that vowel contrasts are easier to discriminate than consonant contrasts, as evidenced by Martinez and colleagues (2008) and Uhler and colleagues (2015;2018), all of which utilize VRISD. Martinez and colleagues (2008) found that infants had variable performance on consonant contrasts regardless of hearing status and type of consonant contrast. Uhler and colleagues (2015) found that 91% of normal hearing infants could discriminate an /a-i/ contrast whereas only 71% of normal hearing infants could discriminate a /ba-da/ contrast (p. 6-7). Moreover, results from this study noted that a higher presentation level was required for infants to successfully discriminate /ba-da/ contrasts.

These results, while limited, have implications in terms of infant auditory development. It appears that the ability to discriminate between vowel sounds may emerge before the ability to discriminate between consonant sounds, at least when using behavioral measures. Pisoni (1973) noted that differences in discrimination between vowels and consonants may be related to differential use of short-term auditory memory for vowels versus consonants (p. 7).

Normal hearing infants versus infants with hearing loss:

A total of 7 papers utilized subjects with hearing loss, including 3 OPP papers and 4 VRISD papers. All VRISD papers with infants with hearing loss in the subject pool (Eisenberg et al., 2004; Martinez et al., 2004; Uhler et al., 2011;2018) utilized phonetic stimuli and assessed speech discrimination ability. Stimuli were similar across papers, with 3 utilizing CV or vowel contrasts and one using vowel-consonant-vowel contrasts. OPP papers which utilized infants with hearing loss assessed localization abilities (Greico-Calub et al., 2008;2012) and reaction time (Dasika et al., 2009). While the OPP papers that utilized children with hearing loss can provide some information on how they perform auditorily, they do not assess speech discrimination ability, suggesting that more research must be done with phonemic stimuli prior to using OPP to assess speech discrimination ability in infants.

From VRISD results, it appears that in children with hearing loss, consonant contrast discrimination harder to master, when compared to vowel contrast discrimination. This is apparent in Eisenberg et al. (2004) and in Uhler et al (2018). Eisenberg and colleagues (2004) found that infants with normal hearing responded consistently to both vowel and consonant contrasts, whereas infants with hearing loss responded consistently only to vowel contrasts. Uhler and colleagues (2018) reported that 95% of infants with hearing loss successfully reached criterion for the /a-i/ contrast while only 50% of infants with hearing loss reached criterion for the /ba-da/ contrast, meaning that the consonant contrast was more difficult to discriminate. The proportion of infants with hearing loss reaching criterion was significantly higher for the /a-i/ contrast versus the /ba-da/ contrast,(50% of HL infants versus 71% of NH infants) , whereas the proportion of infants with normal hearing did not have a significant difference between type of contrast (Uhler et al., 2018, p. 365).

It is difficult to compare results between OPP and VRISD studies as stimuli and dependent variables do not match. One potential clinical benefit of OPP is that it can be used to measure a variety of auditory capabilities. While this is beneficial, the aim of this review is to evaluate clinical utility of speech discrimination ability. Considering that no selected OPP study has assessed speech discrimination ability in infants with hearing loss, more research will be needed prior to its potential use as validation of amplification fittings.

While results from VRISD papers are limited, they suggest that infants with hearing loss, much like infants with normal hearing, struggle to discriminate consonants, while vowel discrimination ability is more like that of infants with normal hearing. This can suggest that infants with hearing loss develop vowel discrimination abilities prior to consonant discrimination abilities, which is what is believed to occur in infants with normal hearing. Alternatively, it could be that vowel contrasts are more salient. Additionally, Uhler and colleagues (2011) noted that vowel and voice onset time discrimination ability appeared to develop prior to place of articulation (consonant) discrimination ability in infants with hearing loss utilizing cochlear implants. This can potentially help a clinician create goals in verifying and validating amplification in infants with hearing loss. By knowing that infants develop vowel discrimination abilities first, it would be appropriate to work on mastering vowel discrimination in an auditory training program prior to mastering consonant discrimination abilities.

Limitations:

An issue found across papers is the limited sample sizes used in the studies reviewed. With an average sample size of 39.3 and 18.8 for OPP and VRISD respectively, larger-scale studies must be performed to generalize results to infants with and without hearing loss. This is crucial before either procedure can be transitioned to the clinic. Moreover, data on infants with hearing loss is particularly limited, with 3 OPP and 4 VRISD studies including children with hearing loss in their subject pool. Considering this, more studies with these infants are required to generalize results. While data on children with hearing loss is limited, general trends within results are noteworthy. For example, Greico-Calub and colleagues (2008) found that some toddlers with bilateral cochlear implants had localization acuity that was on par with their normal hearing peers. Moreover, Uhler and colleagues (2011) found that 2 out of 3 infants with cochlear implants performed similarly to their normal hearing peers on at least 3 out of 5 speech sound contrasts. It would be optimal to have normative data as a function of age and hearing loss so that clinical results can be more easily interpreted.

Of studies that utilized earphones rather than soundfield presentation, only the right ear was used during testing. No VRISD studies utilized headphones. Studies with ear specific information are warranted prior to implementing procedures clinically. This includes studies performed with headphones with normal hearing children and unilaterally aided studies in children with hearing loss.

Clinical Utility:

A goal of this paper is to evaluate the clinical utility of both reviewed procedures. There are modifications that need to be made to procedures that will allow for them to be used more efficiently in clinical practice. These modifications would also need to be studied in a sufficiently large population of infants with and without hearing loss prior to implementing them universally in a clinical setting. Stimulus parameters, such as contrasts and starting intensity levels, must be refined and researched in depth. In terms of using behavioral speech discrimination measures as an outcome measure for amplification fittings, stimuli should be standardized and tailored to goals of amplification and developmental milestones. For example, it is noted that children with high-frequency hearing losses often have trouble discriminating high frequency sounds such as /s/, /sh/, or /f/ (Stelmachowicz et al., 2001, p. 2188). While some reviewed studies utilized /s/ and /sh/ stimuli within different contexts, only Eilers and colleagues' 1979 study directly assessed a contrast with these two stimuli. Results from this study indicate that infants with normal hearing can discriminate between /sa/ and /sha/, however this contrast has not been utilized in any study with infants with hearing loss.

Another use of a behavioral speech discrimination measure for infants is assessing speech discrimination in noise. Nozza and colleagues (1991b) found that infants aged 7-11 months could successfully discriminate a /ba-ga/ contrast utilizing VRISD with a group mean of 46.3 dB SPL in 48 dB SPL of noise. Leibold and colleagues (2016) assessed speech detection ability with two-talker babble and speech-shaped noise masking conditions and found that infants could successfully detect speech in either condition utilizing OPP. It is pertinent to note that Leibold and colleagues' 2016 study does not assess speech discrimination ability. Considering that data on speech discrimination in noise is limited or non-existent for both procedures, more research is warranted prior to implementing a behavioral speech in noise task for infants clinically.

A possible issue in implementing these procedures clinically is the ability to train clinicians to use either procedure. Olsho and colleagues (1987) noted that training to observe for OPP testing takes approximately one month (p. 621). In clinical practice, this is not feasible, especially for practices that split their time among pediatric and adult populations. In order to utilize OPP clinically, the training time for testers must be drastically reduced. Training for VRISD testing is not discussed in depth in terms of the tester/experimenter, but the procedure is like that of visual reinforcement audiometry and is more dependent on the child being conditioned to complete the procedure.

Another issue that will need to be addressed prior to the clinical implementation of VRISD or OPP is that of keeping experimenters blinded during testing. Realistically, blinding testers and assistants during testing can prolong the shaping, training, and testing of infants which can lead to additional visits required to complete testing. Unless software is commercially available and accounts for blinding of the tester and assistant, it may be more feasible to conduct these procedures in a clinical setting without the use of blinding. This would have the same limitations as clinical VRA practices, in that it will not be bias-free. However, if there is repeatability in discrimination across runs, confidence in the results would be increased.

An issue that must be addressed prior to implementing these procedures clinically is the implementation of universal stimuli used across clinics. A variety of stimuli were used in the reviewed studies, which has been discussed in depth in the results section of this paper. One set of stimuli that is commercially available for speech detection procedures are the Ling 6 sounds. These sounds, /a, i, u, m, s, sh/ are commonly used in hearing aid listening checks and encompass a variety of frequencies across the speech spectrum. Alternatively, it might be useful to include easier and more difficulty vowel and consonant contrasts, similar to the approach used in VRASPAC. This would allow for a discrimination task between vowel-vowel contrasts, consonant-vowel contrasts, and consonant-consonant contrasts. One caveat to this approach would be that contrasts would only include phonemes in isolation rather than consonant-vowel clusters, which have been studied more in depth.

OPP would not be able to be utilized through this proposed procedure because the observer must be blinded to whether a change trial occurred. Considering this and the long training time to learn to observe for the procedure, VRISD appears to be more feasible for a clinical setting. There is more literature on VRISD than OPP used in children with and without hearing loss, evaluation is faster and more efficient, and the procedure can be used as an outcome measure in children with hearing loss. A potential downside is that some infants might respond in ways other than a head turn, which OPP handles better than VRISD. In terms of stimuli, it will be critical to select contrasts that are developmentally appropriate and/or that are of particular importance to the habilitation needs of a particular child. For example, one can use a /sh-s/ contrast on an infant with high-frequency hearing loss while aided to determine if a child can perceive the difference between the two high frequency phonemes with amplification. A contrast such as /sh-s/ would be difficult to discriminate for any infant, with or without hearing loss, and therefore it would only be developmentally appropriate to utilize this contrast in infants that already have vowel discrimination ability. Considering this, a clinician might set speech sound discrimination goals for infants with hearing loss to follow a developmental timeline. If a infant with hearing loss cannot discriminate vowels, it would be usually be unrealistic to set a short term goal of mastering consonant discrimination. Moreover, results from aided behavioral speech discrimination testing could potentially be used as a counseling tool for parents and caregivers of infants with hearing loss.

A proposed procedure for testing in clinical populations based off VRISD is shown below. Research on the procedure using large numbers of children with varied hearing status would be needed before implementation.

Proposed Clinical Procedure: modified from VRISD

1. Have the infant enter the room with caregiver while background stimulus is playing via soundfield at 50 dB SPL (ie: /s/)
    a. Infants with hearing loss should be wearing their amplification in the booth if evaluating the aided listening condition.
2. Once the infant appears to habituate to sound, tester should present target stimulus at a level around 20 dB higher than background stimulus (70 dB SPL), paired simultaneously with visual reinforcement (ie: /s/ /s/ /s/ /s/ /SH/ /SH/ /SH/ /s/ /s/ /s/; sh used as target)

3. Once the infant begins to make anticipatory head turns to a change in stimulus, lower target stimulus by 10 dB and repeat until the infant begins to make anticipatory head turns at this level.
4. Lower target level to that of background stimulus level, present at this level in a way that is not patterned, to avoid false positives from the patient. Target trials should be presented at least 50% of the time. A head turn response must be observed within 4 seconds of target stimulus presentation.
5. Initiate a probe trial with the target stimulus presented 20 dB above the background stimulus if the infant appears inattentive.
6. Present target stimulus 10-20 times, score results to obtain a percentage.
    a. Track scores over time as an amplification outcome measure, if desired.
7. Repeat steps 1-6 with other stimuli if desired.

This proposed procedure essentially follows a standard clinical visual reinforcement audiometry (VRA) procedure. Some differences are that with VRA, there would not be a repeating background stimulus as the child enters the room and the child is conditioned to turn their head to a change in sound, rather than a presence of sound. This proposed procedure would be simple to implement in clinical practice for audiologists that are experienced in using VRA with their patients. Benefits of this proposed procedure are that the procedure can be used as an objective outcome for amplification, as a counseling tool for parents, and that the proposed procedure does not require extensive training to conduct. A drawback of this proposed procedure is that the tester is not blinded, which can bias scoring. Moreover, the tester must be able to conduct the task at a quick pace, as infants and toddlers quickly habituate and fatigue during visual reinforcement testing. This proposed procedure has not been used clinically or in research and would need to be researched in depth prior to its implementation. Another option for clinical implementation for VRISD is the use of Intelligent Hearing Systems iVRA system, which is an automated visual reinforcement audiometry system that also has a VRISD paradigm built in (Intelligent Hearing Systems, n.d.). This is commercially available however the iVRISD module has not been normed on infants. The use of an automated system to run testing can allow for testing to be conducted with one clinician present and could possibly utilized to address the issue of blinding the clinician to trial type. The VRASPAC approach could also potentially be standardized and made commercially available.

Overall, while results from research are promising, more research must be conducted with larger sample sizes prior to implementation of a behavioral speech discrimination procedure in infants. 5 out of 9 included OPP studies were conducted at or in affiliation with the Department of Speech and Hearing Sciences, University of Washington. Therefore, additional studies should be conducted at additional sites in attempt to replicate their results (Werner et al., 1992; Greico-Calub et al., 2008; Dasika et al, 2009; Lalonde & Werner, 2019; Walker et al., 2019). The need for larger sample sizes holds for both OPP and VRISD. Each of these approaches had strengths and weaknesses, but in general the transition of VRISD to the clinic would be more straightforward.

CONCLUSION

In this systematic review, existing literature on the observer-based psychoacoustic procedure (OPP) and visual reinforcement infant speech discrimination (VRISD) were reviewed in terms of potential utility in a standard clinical setting, differences in results from normal hearing children and children with hearing loss, and changes needed to be made to use these tests in a clinical setting. Infants perform significantly worse than adults on these tasks, and infants with hearing loss appear to struggle more with consonant contrasts compared to their normal hearing peers. Results have shown that infants with and without hearing loss can be assessed with these procedures, however further research is needed to establish an age and hearing loss normed testing protocol in a clinical setting. Several modifications of VRISD have been proposed for clinical application. If implemented clinically, behavioral speech discrimination measures could potentially be used to monitor changes in speech discrimination with development, amplification, and/or aural habilitation.

REFERENCES

Ashmead, D., & Tharpe, A. (1998). A longitudinal investigation of infant auditory sensitivity. *Infant Behavior and Development, 21*, 271. https://doi.org/10.1016/s0163-6383(98)91484-9

Auditec, Inc. (n.d.). Nuchips – Auditec, Inc. https://auditec.com/tag/nuchips/

Bargones, J. Y., & Werner, L. A. (1994). Adults Listen Selectively; Infants Do Not. Psychological Science (0956-7976), 5(3), 170–174.

Bull, D., Eilers, R. E., & Oller, D. K. (1984). Infants' discrimination of intensity variation in multisyllabic stimuli. The Journal of the Acoustical Society of America, 76(1), 13-17. https://doi.org/10.1121/1.391110

Bull, D., Eilers, R. E., & Oller, D. K. (1985). Infants' discrimination of final syllable fundamental frequency in multisyllabic stimuli. The Journal of the Acoustical Society of America, 77(1), 289-295. https://doi.org/10.1121/1.392270

Cone, B. K. (2015). Infant cortical electrophysiology and perception of vowel contrasts. International Journal of Psychophysiology, 95(2), 65-76. https://doi.org/10.1016/j.ijpsycho.2014.06.002

Cone, B., & Whitaker, R. (2013). Dynamics of infant cortical auditory evoked potentials (CAEPs) for tone and speech tokens. International Journal of Pediatric Otorhinolaryngology, 77(7), 1162-1173. https://doi.org/10.1016/j.ijporl.2013.04.030

Dasika, V. K., Werner, L. A., Norton, S. J., Nie, K., & Rubinstein, J. T. (2009). Measuring sound detection and reaction time in infant and toddler cochlear implant recipients using an observer-based procedure: A first report. Ear & Hearing, 30(2), 250-261. https://doi.org/10.1097/aud.0b013e3181986dfe

Dickerson, D., Swafford, M., & Morrow-Almeida, H. (2017). Early hearing detection and intervention. Underlying Standards that Support Population Health Improvement, 19-47. https://doi.org/10.4324/9781315153230-2

Eilers, R. E., Gavin, W., & Wilson, W. R. (1979). Linguistic experience and phonemic perception in infancy: A Crosslinguistic study. Child Development, 50(1), 14. https://doi.org/10.2307/1129035

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. https://doi.org/10.1044/jshr.2004.766

Eisenberg, L. S., Martinez, A. S., & Boothroyd, A. (2004). Perception of phonetic contrasts in infants: Development of the VRASPAC. International Congress Series, 1273, 364-367. https://doi.org/10.1016/j.ics.2004.09.001

Eisenberg, L. S., Martinez, A. S., & Boothroyd, A. (2007). Assessing auditory capabilities in young children. International Journal of Pediatric Otorhinolaryngology, 71(9), 1339-1350. https://doi.org/10.1016/j.ijporl.2007.05.017

Goodsitt, J. V., Morse, P. A., Hoeve, J. N., & Cowan, N. (1984). Infant speech recognition in Multisyllabic contexts. Child Development, 55(3), 903. https://doi.org/10.2307/1130141

Grieco-Calub, T. M., Litovsky, R. Y., & Werner, L. A. (2008). Using the observer-based psychophysical procedure to assess localization acuity in toddlers who use bilateral cochlear implants. Otology & Neurotology, 29(2), 235-239. https://doi.org/10.1097/mao.0b013e31816250fe

Grieco-Calub, T. M., & Litovsky, R. Y. (2012). Spatial acuity in 2-to-3-Year-Old children with normal acoustic hearing, unilateral cochlear implants, and bilateral cochlear implants. Ear & Hearing, 33(5), 561-572. https://doi.org/10.1097/aud.0b013e31824c7801

Hillenbrand, J. (1984). Speech perception by infants: Categorization based on nasal consonant place of articulation. The Journal of the Acoustical Society of America, 75(5), 1613-1622. https://doi.org/10.1121/1.390871

Hillenbrand, J., Minifie, F. D., & Edwards, T. J. (1979). Tempo of spectrum change as a cue in speech-sound discrimination by infants. Journal of Speech, Language, and Hearing Research, 22(1), 147-165. https://doi.org/10.1044/jshr.2201.147

JCIH year 2000 position statement: Principles and guidelines for early hearing detection and intervention programs. (2000). https://doi.org/10.1044/policy.ps2000-00035

Lalonde, K., & Werner, L. A. (2019). Infants and adults use visual cues to improve detection and discrimination of speech in noise. Journal of Speech, Language, and Hearing Research, 62(10), 3860-3875. https://doi.org/10.1044/2019_jslhr-h-19-0106

Leibold, L. J., Yarnell Bonino, A., & Buss, E. (2016). Masked speech perception thresholds in infants, children, and adults. Ear & Hearing, 37(3), 345-353. https://doi.org/10.1097/aud.0000000000000270

Lieu, J. E., Kenna, M., Anne, S., & Davidson, L. (2020). Hearing loss in children. JAMA, 324(21), 2195. https://doi.org/10.1001/jama.2020.17647

Martinez, A., Eisenberg, L., Boothroyd, A., & Visser-Dumont, L. (2008). Assessing speech pattern contrast perception in infants. Otology & Neurotology, 29(2), 183-188. https://doi.org/10.1097/mao.0b013e3181625114

Moore, J. M., Thompson, G., & Thompson, M. (1975). Auditory localization of infants as a function of reinforcement conditions. *Journal of Speech and Hearing Disorders*, *40*(1), 29-34. https://doi.org/10.1044/jshd.4001.29

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, Language, and Hearing Research, 34(3), 643-650. https://doi.org/10.1044/jshr.3403.643

Nozza, R. J., Rossman, R. N., & Bond, L. C. (1991). Infant-adult differences in unmasked thresholds for the discrimination of consonant-vowel syllable pairs. International Journal of Audiology, 30(2), 102-112. https://doi.org/10.3109/00206099109072875

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(5), 627-640. https://doi.org/10.1037/0012-1649.23.5.627

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception & Psychophysics, 13(2), 253-260. https://doi.org/10.3758/bf03214136

Pourarian, S., Khademi, B., Pishva, N., & Jamali, A. (2012). Prevalence of hearing loss in newborns admitted to neonatal intensive care unit. Iranian journal of otorhinolaryngology, 24(68), 129–134.

Sinnott, J. M., & Aslin, R. N. (1985). Frequency and intensity discrimination in human infants and adults. The Journal of the Acoustical Society of America, 78(6), 1986-1992. https://doi.org/10.1121/1.392655

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., & Lewis, D. E. (2001). Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults.

The Journal of the Acoustical Society of America, 110(4), 2183–2190. https://doi-org.ezproxy.gc.cuny.edu/10.1121/1.1400757

Teller, D. Y. (1979). The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. *Infant Behavior and Development, 2*, 135-153. https://doi.org/10.1016/s0163-6383(79)80016-8

Trehub, S. E., Schneider, B. A., Thorpe, L. A., & Judge, P. (1991). Observational measures of auditory sensitivity in early infancy. *Developmental Psychology*, *27*(1), 40-49. https://doi.org/10.1037/0012-1649.27.1.40

Uhler, K. M., Baca, R., Dudas, E., & Fredrickson, T. (2015). Refining stimulus parameters in assessing infant speech perception using visual reinforcement infant speech discrimination: Sensation level. Journal of the American Academy of Audiology, 26(10), 807-814. https://doi.org/10.3766/jaaa.14093

Uhler, K. M., Gifford, R. H., Forster, J. E., Anderson, M., Tierney, E., Claycomb, S. D., & Werner, L. A. (2018). Refining stimulus parameters in assessing infant speech perception using visual reinforcement infant speech discrimination in infants with and without hearing loss: Presentation level. Journal of the American Academy of Audiology, 29(09), 847-854. https://doi.org/10.3766/jaaa.17061

Uhler, K., Yoshinaga-Itano, C., Gabbard, S. A., Rothpletz, A. M., & Jenkins, H. (2011). Longitudinal infant speech perception in young cochlear implant users. Journal of the American Academy of Audiology, 22(03), 129-142. https://doi.org/10.3766/jaaa.22.3.2

Walker, B. A., Gerhards, C. M., Werner, L. A., & Horn, D. L. (2019). Amplitude modulation detection and temporal modulation cutoff frequency in normal hearing infants. The

Journal of the Acoustical Society of America, 145(6), 3667-3674. https://doi.org/10.1121/1.5111757

Werner, L. A., Marean, G. C., Halpin, C. F., Spetner, N. B., & Gillenwater, J. M. (1992). Infant auditory temporal acuity: Gap detection. Child Development, 63(2), 260. https://doi.org/10.2307/1131477

Widen, Judith E.; O'Grady, Gwendolyn M. Using visual reinforcement audiometry in the assessment of hearing in infants, The Hearing Journal: November 2002 - Volume 55 - Issue 11 - p 28-36 doi: 10.1097/01.HJ.0000324169.66607.69

Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. (2007). https://doi.org/10.1044/policy.ps2007-00281