ASSESSING INFANT SPEECH DISCRIMINATION USING PUPILLARY DILATION RESPONSE

by

Sam Heilenbach

A THESIS

Presented to the Department of Psychology
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

August 2024

An Abstract of the Thesis of

Sam Heilenbach for the degree of Bachelor of Science
in the Department of Psychology to be taken August 2024

Title: Assessing Infant Speech Discrimination Using Pupillary Dilation Response

Approved: *Dare Baldwin, Ph.D.*
Primary Thesis Advisor

Perceptual narrowing is central to the ability of infants in acquiring language competency. The broad sensitivity seen in younger infants to even the smallest units of speech – phonemes – becomes fine-tuned toward the end of their first year of life to the specific phonemes of their native language(s), enabling them to partition continuous speech into discrete words, essential for acquiring language proficiency. Perceptual narrowing refers to the gradual reduction in sensitivity to phonetic contrasts that are not relevant in the native language, along with an enhanced ability to recognize and distinguish phonetic contrasts that are meaningful in the native language. Understanding the timing and mechanisms of this process is key to understanding speech and language development more broadly. Current methods for studying perceptual narrowing rely on behavioral methods that require lengthy training periods, which result in high rates of attrition and may resultingly limit the generalizability of these findings. This study investigated the potential of sound-induced pupil response (SIPR) to measure infant speech-sound discrimination. We hypothesized that SIPR will accurately index auditory discrimination changes as perceptual narrowing unfolds. Using a mixed factorial-design, infants aged 10-12 months were exposed to native and non-native phonetic contrasts while their pupil responses were measured. Our sample of pilot data provides clear evidence validating SIPR as a sensitive

measure of infant response to language. However, the sample is as yet too small to offer conclusive findings regarding perceptual narrowing. In any case, the SIPR holds great promise as a new, highly sensitive tool for investigating language development in early infancy. Thus, the significance of this research lies in its potential to enhance our understanding of developmental change in early auditory discrimination abilities.

Acknowledgements

Firstly, my thanks to the University of Oregon's Undergraduate Research Opportunity Program (UROP) and specifically the Vice Provost of Research and Innovation Undergraduate Fellowship for providing me the funding necessary to support my full-time involvement with this project over the course of summer 2023.

I would also like to thank Dr. Dare Baldwin and the entire Acquiring Minds Lab, both past and present! Without Dr. Baldwin's continued support, guidance, and involvement, my thesis would not be what it is today. Further, without the extraordinary work that Dr. Baldwin and colleagues in the Acquiring Minds Lab have conducted for the past few decades, the methods and questions addressed in my thesis would not be where they are.

Finally, I would like to extend my sincerest gratitude to Dr. Avinash Singh Bala and his colleagues in the Baby Hearing Lab. Without Dr. Bala's guidance and input throughout my thesis process, my involvement with this project would not be what it is. Further, thank you to Dr. Bala, Perceptivo LLC, and the members of his lab, both past and present, for their wonderful work developing these ground-breaking methods and technologies.

Table of Contents

List of Figures

List of Tables

Introduction

A six-month-old sits in their highchair, gazing attentively at their caregiver across the table. Their caregiver's mouth opens, flexes, and contorts. The child is awash in a stream of continuous sounds – both speech from their caregiver and the clamor of daily life. The caregiver says, “Can you say cat?”. For English language-proficient adults, this sentence contains readily interpretable linguistic content. However, to a six-month-old primed to learn any of the world's languages (Werker et al., 2012), the meaning of these utterances is not so clearly mapped onto semantic meaning.

How babies acquire language has inspired rigorous research from diverse academic disciplines. Over the past few decades, contributions from linguists, psychologists, and experts in computational modeling have aided in elucidating the mechanisms underlying language acquisition and speech perception. A shared interest in how infants decode spoken language unites these researchers from disparate fields. Our understanding of speech perception and its underlying mechanisms have undoubtedly benefited from this confluence of multidisciplinary research. However, given certain methodological limitations (See Literature Review below) of previous methods, there are grounds to believe that much is still unknown.

Recent advancements reveal new ways to explore these questions. Increasing sophistication and accessibility of technical tools enable researchers to glean a more complete picture of how infants learn language. In this study, we explore the potential of sound-induced pupil response (SIPR), a specific application of pupillometry, to measure the ability of infants to discriminate between speech sounds—a critical facet of language development.

Literature Review

Speech Perception & Language Acquisition

Central to this body of research is the phenomenon of perceptual narrowing. Perceptual narrowing is the process by which infants initially broad sensitivity to speech sounds, or phonemes, becomes fine-tuned to the specific sounds of their native language(s). In other words, infants are born with the capacity to differentiate between speech sounds from any language, but through perceptual narrowing, they become experts at attending to only those contrasts that are meaningful within their linguistic environment (Werker et al., 2012; Kuhl, 2004).

Prior research indicates a two-month sensitive period for perceptual narrowing between 8-10 months (Kuhl, 2004). Thus, a monolingual baby younger than 8 months growing up in a Japanese-speaking household would likely still attend to the phonetic difference between phonemes /la/ and /ra/. Yet, by the age of 10 months, the same baby would no longer distinguish between the phonemes /la/ and /ra/, because they are in the same phonetic category in Japanese, therefore, are not distinguished by Japanese speakers(Kuhl, 2004).

Behavioral methods for studying the perceptual narrowing process, such as the head-turn preference procedure, have outlined the developmental trajectory of this phenomenon, revealing how infants gradually lose sensitivity to non-native phonetic contrasts while enhancing their responsiveness to native contrasts (Kuhl, 2004; Gervain & Werker, 2008). Perceptual attunement to the speech sounds in their native language allows infants to partition continuous streams of speech into discrete words, making it a crucial early step in language acquisition (Kuhl, 2004; Saffran et al., 1996; Werker et al., 2012).

These approaches have facilitated the exploration of the role of learning mechanisms, such as the role of prosodic cues and distributional learning, in driving perceptual narrowing

(Kuhl, 2004). Indeed, once infants can reliably identify a speech sound as a phoneme belonging to a particular phonetic category in their language, they can start to infer where one word ends and the next begins. Statistical regularities within the lexicon of a given language, regarding what phonemes are likely to be combined to form meaningful utterances, have been identified as an important mechanism by which infants use their ability to distinguish phonemes to progress on their language learning journey (Saffran et al., 1996; Saffran & Kirkham, 2018; Werker et al., 2012). Despite the significant insights gained, prior approaches to studying perceptual narrowing have limitations. For example, evidence for perceptual narrowing has been largely limited to behavioral, head-turn techniques or looking time paradigms(Gervain & Mehler, 2010).

In one popular variant of the head-turn procedure, researchers repeatedly present the infants with a ‘prototype vowel’ that represents a standard variant of a vowel phoneme in English. Before the experiment begins, babies are trained to respond to sounds by turning their heads in a certain direction, and this behavior is positively reinforced. During experimentation, at random intervals, the infants are then presented with a different prototype vowel representing a variant of a prototypical Swedish vowel phoneme (Kuhl, 2004). Infants that successfully discriminate between two prototype vowel speech sounds are conditioned to turn their heads away from the research assistant. As in the training phase, they are rewarded for successfully doing so by briefly viewing an animated stuffed animal or similar toy (Kuhl, 2004; See Figure 1). If the participants who complete the training systematically differ from those who do not, this attrition could significantly impact the generalizability of the results. For example, suppose the infants who drop out of these studies tend to have limited attentional capabilities. In that case, the findings might not accurately reflect the developmental trajectory of perceptual narrowing in the broader infant population. Thus, tasks that rely on behavioral indicators, such as the head-turn task, are limited in generalizability because many infants cannot learn the behavioral responses proficiently enough for experimentation. The implications of these limitations in generalizability are revisited in the Discussion section.

Figure 1 – Head-Turn Task

a) In the left-most panel, an infant looking towards the research assistant during the head turn procedure, and b) the right-most panel, an infant turning their head away from the assistant in the right frame, indicating that they heard a change in the prototype vowel that they were being played (Kuhl 2004).

As with all areas of academic research, the study of infant speech perception can be revolutionized by the emergence of increasingly economical, accessible, and advanced technologies. In the following section, we will explore the principles of pupillometry, its application in language development, and its potential to serve as a measure of processing of speech sounds in infants, thereby helping us understand early language acquisition.

Pupillometry and the Sound-Induced Pupillary Response (SIPR)

Pupillometry – or measuring pupil dilation across time – offers a novel and potentially more usable approach to indexing speech sound processing of infants. It could allow researchers to assay the progression infants undertake through the perceptual narrowing process more accurately and with more generalizability. Below, I review the history of the development and use of the pupillometry technique in psychological research and then explain its application to measuring speech perception in infants.

Seminal cognitive scientists Kahneman and Beatty (1966) found that pupil dilation responses (PDRs) reflect various types of cognitive processing. They identified two distinct forms of PDRs. Tonic PDRs are characterized by a gradual pupil dilation as the demands on a person's memory capacity increase and by rapid constriction once the cognitive effort has subsided (Kahneman & Beatty, 1966). Recently, researchers have become interested in phasic PDRs. Phasic PDRs often occur as part of the orienting reflex (OR), a central component of covert orienting aspects of attention that direct our attention to novel or potentially harmful stimuli (Sokolovc, 1990). This reflex is particularly significant in auditory perception, as unexpected or novel sounds frequently trigger the OR, engaging attentional processes to assess the new environmental stimuli. Unlike tonic PDRs, phasic PDRs occur on a shorter timescale and respond to different environmental stimuli, making them a potentially rich source of insight into involuntary sensory processes like those implicated in the orienting response (Bala et al., 2020).

The foundational understanding of the relation between PDRs and cognitive processing, first enumerated by Kahneman & Beatty (1966), has quickly sparked innovative new methods in many psychological subdisciplines (Sirois & Brisson, 2014). In particular, the relationship between phasic PDRs and the orienting reflex highlights its suitability for studying auditory detection and discrimination in adults and babies alike. Indeed, Bala and colleagues (2020) demonstrated that measuring sound-induced pupillary responses (SIPRs) can provide as sensitive a measure of adult auditory detection as methods that rely on voluntary reporting. They also found that SIPRs provide a sensitive index of adults' auditory discrimination capabilities using a habituation and recovery paradigm (Bala et al., 2020).

Further investigations by Bala and colleagues included a study on 37 infants aged 6–12 months, showing that SIPRs in infants were reliable measures of auditory capabilities in infants (Bala et al., NWAVRM; Bala et al., 2020). In their experimental paradigm, the attention of infant participants was engaged using a simple yet captivating animation, made with the consideration of Kidd and colleagues' realization that infants attend to visual streams that are neither too complex nor too simple, which proved effective in maintaining focus (Kidd et al., 2012). This study employed an auditory sensitivity task, where some trials contained no sound stimulus, and others comprised sound stimuli of various intensities. Their findings revealed that sound-elicited PDRs habituate similarly in infants as they do in adults, suggesting that a PDR-based assay could be a promising method for assessing auditory discrimination in infants using an oddball or habituation and recovery paradigm (Bala et al., 2020). Thus, the data demonstrate that, just as in adults, SIPRs can provide a sensitive measure of auditory detection and discrimination in infants.

In the present study, both participants were in the 10-12-month-old age range. Prior research on the normative developmental timescale of perceptual narrowing has found that after ten months of age, infants no longer attend to phoneme contrasts that are not native to the language(s) they are emersed in (Kuhl, 2004). We expect that our participants, being in the 10-12-month-old group will demonstrate PDRs indicative of auditory discrimination only in response to English native phoneme contrasts (ka vs. ga) since they are further along in the perceptual narrowing process and, therefore, only attend to phoneme contrasts that are distinguished in their native language.

We exploratorily predict that the PDR measures of auditory discrimination in infants used in this study might offer a more sensitive and generalizable gauge of discrimination capabilities of infants and their advancement through the perceptual narrowing process than previously collected behavioral methods.

Methods

Participants

Methods and recruitment procedures used in this study were approved by the Institutional Review Board of the University of Oregon. I participated in the testing of nine infants, of which two were selected for detailed analysis. The two participants selected for further data analysis were 11 months old. The full-scale study will comprise 45-50 infants in two age categories, 6- 8 months of age and 10-12 months of age. Participants were recruited using the University of Oregon Team Duckling Developmental Database (https://teamduckling.uoregon.edu/). Data were excluded from analysis when infant participants could not complete a full batch of 30 trials (totaling 3 minutes) due to fussiness.

Testing Environment & Equipment

Trials occurred in an anechoic sound-proof booth (Whisper Room Inc.'s SE 2000 series). The booth's ambient sound and light levels were recorded throughout the data collection period to control for background noise that would impact participants' pupil reactivity. Infants were seated either in a car seat or, if necessary, in their caregivers' lap, and positioned approximately 48 centimeters from a monitor/camera/infrared light mount. The mount was made up of two infrared cameras, two infrared light arrays, a monitor, a Raspberry Pi, and a LabJack (U3-LV) to coordinate stimulus presentation and onset of recording.

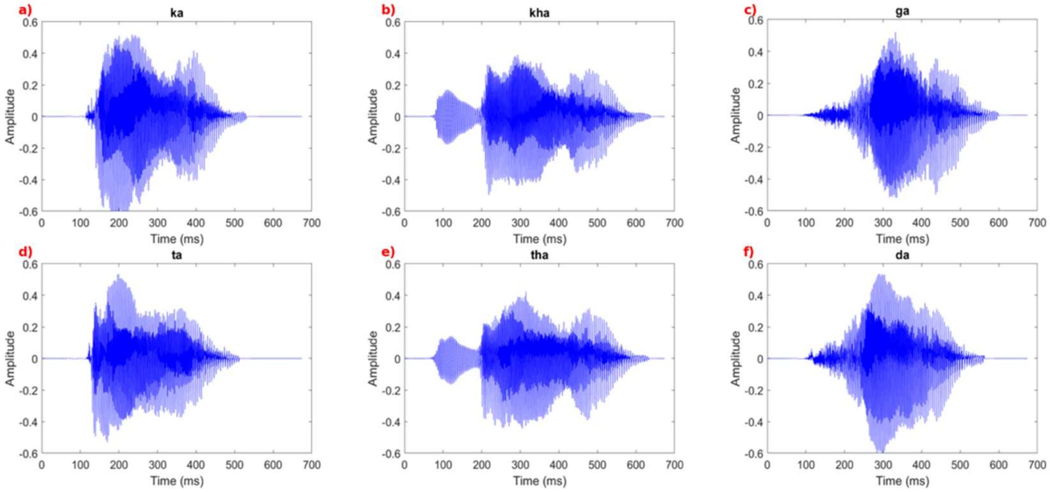


Figure 2 – Time/Amplitude Graphs for Sound Stimuli

Figure 2: a-c) These are the time/amplitude plots for the primary stimulus set where ka represents the habituation stimulus, kha represents the English non-native contrast oddball, and ga represents the English native contrast oddball stimulus. e-f) The time domain graphs for the alternate set of phoneme stimuli where ta is the habituation stimulus, that is the English non-native contrast oddball, and da is the English native contrast oddball. These sound stimuli were equalized, concerning duration and overall amplitude, to the extent possible while retaining the acoustic qualities (e.g., pitch and inflection) that distinguish phonemes in natural language.

Participants were presented with 30 total phonemes in each 3-minute batch. Sounds were generated using the Google Text-to-Speech (gTTS) library in Python. We used computer-generated sound stimuli instead of human-produced speech to ensure the quality of the data collected. Computer-generated stimuli allow precise control over the variables that differentiate phonemes in natural language, such as tone and inflection. Human-produced recordings could introduce subtle differences in amplitude and duration, potentially confounding the discrimination study results by providing unintended cues. Figure 2 displays time-domain graphs of each sound stimulus, adjusted in Audacity – a digital sound editing tool – to have similar amplitudes (within 1 dB) and durations, minimizing any acoustic cues that might confound phoneme discrimination results.

Each batch of trials featured a habituation sound (e.g., /ka/) and two oddball sounds (e.g., /kha/ or /ga/). The phonemes /ka/, /kha/, and /ga/ were chosen as a primary set of trial stimuli (representing an English native contrast between /ka/ and /ga/ and an English non-native, but Hindi native, contrast of /ka/ and /kha/). An alternate set of phonemes was also created: /ta/, /da/ (English native contrast), and /tha/ (English non-native, but Hindi native, contrast) were selected as an alternate set of trial stimuli. These specific phonemes were chosen based on their amenability to sound calibration and resemblance to the equivalent, naturally generated phonemes. Sound stimuli were calibrated to maintain consistent sound levels across different phoneme audio stimuli, ensuring that the magnitude of the PDRs was not influenced by variations in sound intensity. Sounds were converted from their digital form for analog presentation by a Cambridge Audio (Dac Magic 100) digital-to-analog sound converter and played through noise-canceling headphones (Carebio Audiometry Headphones TDH39 DD45) if the infant permitted it or through a speaker (Tang Band W3-1878) if they did not. Sounds were presented within one dBA (dB SPL_A) of 42 dBA for all trials (measured with Bruel and Kjaer Type 2235 SPL meter with a .5-in. microphone).

To keep the pupils of the infants in view during the sound presentation, an attention-grabbing animation was displayed to draw their gaze toward the camera. The attention-grabbing animation was created using the Pygame library in Python.

Experimental Design

Our study utilized a mixed factorial design to investigate developmental changes in how infants process speech sounds, specifically examining the emergence of auditory discrimination capabilities across different age groups. Participants were presented with 30 phoneme sound presentations, or trials, in each batch – with batches lasting 3 minutes.

Participants were divided into two age groups, 6-8 months and 10-12 months, to assess developmental differences in speech sound processing cross-sectionally. This allowed us to explore how perceptual narrowing progresses over these crucial months of infancy. In each batch of 30 trials, infants were presented with two types of phonetic contrasts: English native (ka vs. ga) and English non-native (ka vs. kha). This within-subject factor enabled us to assess the discrimination abilities of each infant across different linguistic contexts within the same batch of trials. This allowed us to study the progression through the perceptual narrowing process for each participant. Phonetic contrast presentation order was pseudo-randomized to control for order effects that could influence performance.

In our study, in each batch of trials, the habituation sound was presented as the sound stimulus for the first five trials. Starting with the sixth trial, one of the oddball phonemes was played every fourth trial.

Stimulus Synchronization

Sound presentation was controlled by custom software (Perceptivo, LLC). Stimulus presentation onset times were recorded and stored by the experimentation software for each of the 30 trials in a given batch. As a control, each time a sound stimulus was played, the experimentation software provided a signal via the LabJack that caused the IR-LED, placed directly in front of each IR camera lens, to illuminate. This allowed us to determine when sounds were presented in the video data from each trial based on the presence or absence of the LED flash.

Procedure

After signing of informed consent, the cameras were pre-focused, and the infant and caregiver were brought into the sound booth. The infant was then seated securely in the car seat or on the caregiver’s lap if they did not tolerate the car seat, and the caregiver was oriented to the camera in the corner of the sound booth, where they could signal researchers if they would like the trial to stop at any point. The caregiver was seated inside the sound booth behind the infant (not directly viewing the monitor) with earplugs to ensure they did not bias their infant’s behavior by cuing the infant to react in a certain way.

Once the cameras had been focused so that the eyes of the participant were in frame and focused, the lights in the sound booth were turned off, the experimentation software was launched, and the batch of trials began.

In each batch, three sounds were played, either through the headphones or a spaker if the participant did not tolerate the headphones. One sound served as the habituation sound and was played more frequently than the two oddball stimuli, which were played only occasionally and served to recover the PDR. At the beginning of each batch, the animation was played five times without a sound playing in tandem; thus, pupil size data from the start of the batch could serve as a valid baseline for pupil size and responsiveness to the visual stimuli, facilitating the interpretation of later sound-induced pupillary responses.

Analysis Plan

Given the unknown distribution of infant PDR data, we opted to use a Signal Detection Theory approach to compare the sensitivity of PDR-based measures of auditory discrimination with behavioral measures such as the head-turn procedure, which yields binary data(i.e., participants either do, or do not, discriminate between the stimuli). Receiver-operating characteristic (ROC) analysis was used to compare population responses to habituation as compared to oddball trials. ROC analysis produces a function, with the area under this function representing the proportion correct, *p*(C) (Bala et al., 2020).. If the PDR data for habituation and test trials have indistinguishable statistical distributions, the *p*(C) would be 0.5, indicating that the participants did not discriminate between habituation and test trials any better than chance alone. However, if the habituation and oddball distributions are robustly distinct, the area under the ROC curve would be larger, and the p(C) would approach 1.0, indicating perfect performance. A *p*(C) of 0.75 is used as a measure for discriminability. Thus, using ROC analysis for sound-induced-pupil-response data allows continuous PDR data to be converted into discrete proportion correct (*p*(C)) data and compared with behavioral measures of discrimination like the head-turn procedure.

Analysis was conducted using custom MATLAB analysis modules. Grayscale video data for each trial, along with a file documenting the time of sound presentation and the specific phoneme stimulus set used for the batch, were automatically saved to a secure hard drive through the experimentation software. The analysis pipeline began by importing the frames from the video of the batch of trials. It then identified the frames that sounds were presented by finding the frames with an illuminated IR-LED; as the IR-LED lit up synchronously with each sound presentation.

Then, a MATLAB script calculated the pupil sizes for all frames where a pupil was in frame utilizing the Circular Hough Transform function. Pupil data for each batch of 30 trials were partitioned into individual segments, one for each trial. Each trial was categorized as either habituation or oddball based on the frame numbers of the video corresponding to that trial. This categorization relied on identifying the frames where the IR-LED was illuminated, which indicates the start of each trial.

Next, for each trial the data were normalized based on the baseline pupil size in the frames leading up to the onset of a given trial's sound presentation. Size data were then run through a low pass filter to reduce the noise introduced into the pupil size measurements based on the imperfection of using pixels to approximate the organic shape of the pupil. Finally, after excluding any trials in which the overall proportion of missing data exceeds 25% or where there were too many consecutive missing data points, the data were visualized.

Data Quality and Preprocessing

The quality of the data collected in this study was ensured through a preprocessing protocol. Each trial consisted of capturing 30 frames prior to and 60 frames after an identifiable flash, with each frame analyzed on a frame-by-frame basis. The data were subjected to black thresholding, and frames with a flash were singled out using the imfindcircles circular Hough transform method in MATLAB. Trials were screened for excessive missing data points, specifically more than 25% missing data or more than three consecutive missing data points (NaNs), which resulted in their exclusion. For those trials that passed this quality check, a Z-score normalization was applied where the baseline pupil size was subtracted (the baseline being the average pupil size during a predefined window encapsulating the trial period). Furthermore, missing data points were interpolated linearly to maintain data integrity.

Results

Data Quality of the Sample

Data for this pilot study comprised 120 trials conducted over four batches of 30 trials, with two infants. Two batches were included from participants AB and AC, respectively. Data from a particular trial was excluded from analysis if MATLAB's imcircles function could not identify a pupil in more than 25% of the frames within that trial or if more than three frames did not include a pupil. Pupil size may not be detected in a frame for several reasons. For instance, if the participant was fussy and not facing the camera, or if they were blinking. Data quality thresholds were determined by identifying the minimum amount of data necessary for the statistical analyses to be valid. Table 1 reports the total count and number of missing trials (collapsed across all trial categories) for all four batches combined, as well as reporting on how many habituations, oddball one, and oddball two trials there were and how many of these trials met the quality criteria described above. Of the 120 total trials run across the four batches from participants AB and AC, 112 of them – or 93% met the quality criteria. This indicates the attention-grabbing animation effectively held the attention of the participants.

Trial Category	**Total Number of Trials**	**Trials that Met Quality Criteria**
Total Trials (all trial types)	120	112 (.93)
Habituation Trials	96	89 (.93)
Oddball_1 Trials	10	9 (.9)
Oddball_2 Trials	14	14 (1.00)

Table 1 – Data Quality Report

General Characteristics of Sound-Induced Pupil Responses (SIPRs)

To validate the use of a pupillometry-adapted oddball paradigm for assessing infant auditory discrimination, it is essential to ensure that participants exhibit a sound-induced pupil response (SIPR). Before determining whether the pupil dilation response (PDR) recovers upon the presentation of oddball stimuli, thus indicating the participant's discrimination of the habituation and oddball stimuli, we must confirm that participants' pupils dilate in response to sound presentation across all trial types.

Given the semi-exploratory nature of this pilot validation study, it is essential to examine the general characteristics of the pupil dilation traces that result from the extraction method outlined in the methods section. First, we plotted all trials that met quality criteria for each batch on the same graph (see Figure 3). Figure 3 includes a plot of all trials (that met quality criteria) for each batch from participants AB and AC.

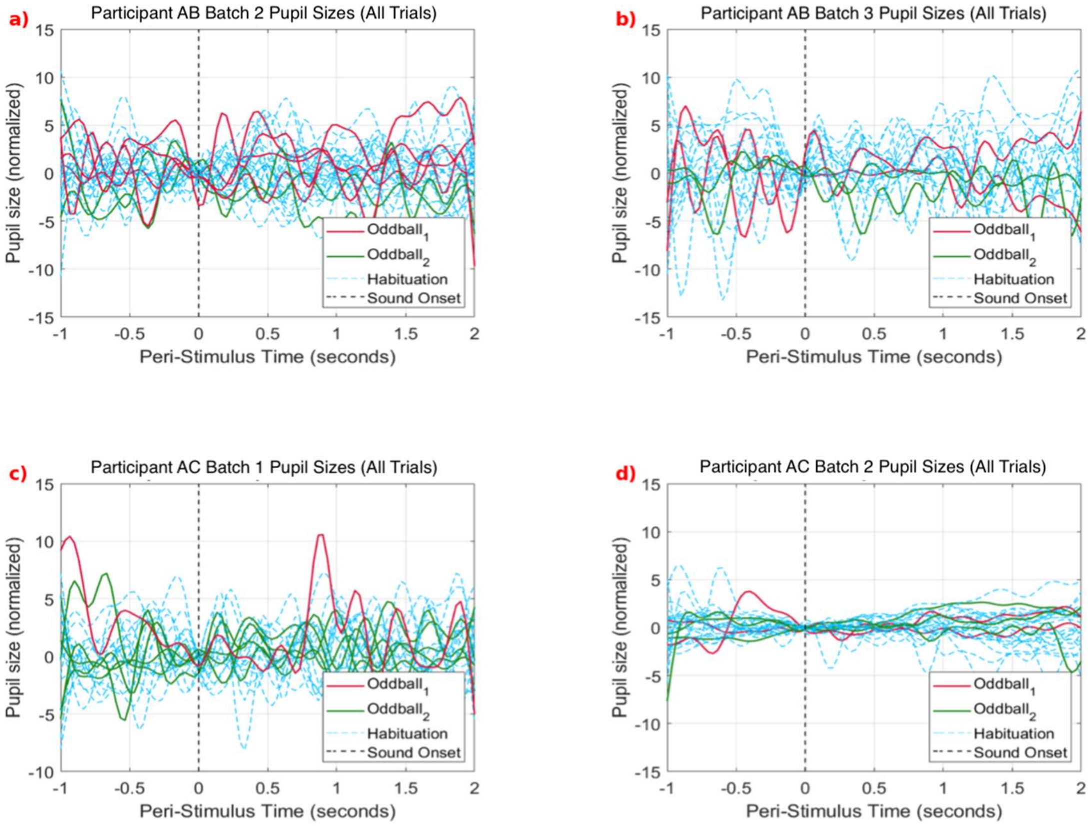
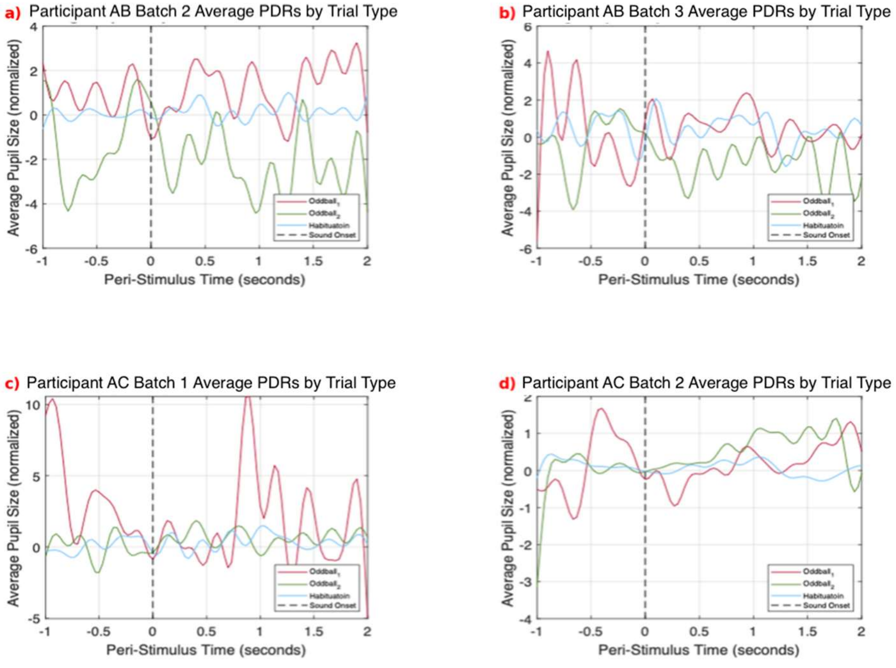


Figure 3 – PDR Traces for All Trials

Pupil sizes from 1 second before sound onset to 2 seconds after sound onset for all trials that met quality criteria. Blue traces indicate habituation trials (/ka/ sound was played), red traces represent oddball1 trials (/ga/ sound was played; English-native contrast), and green traces represent oddball2 trials (/kha/ was played; non-native contrast in English). Subfigures are: a) All trials from participant AB's first batch, b) AB's second batch, c) All trials from participant AC's first batch, d) AC's second batch.

Next, we visualized the average PDR trace for habituation trials, oddball one trials (where the phoneme representing an English-native contrast was presented), and oddball two trials (where the phoneme representing an English non-native contrast was presented) (see Figure 4). Importantly, as was mentioned in the methods, the pupil size for each trial was normalized by subtracting a dynamic baseline pupil size (computed based on the pupil size in the window from .25 seconds before up until .25 seconds after sound presentation). This ensures that pupil sizes would indicate dilation (an increase in pupil size after sound onset) or constriction (a decrease in pupil size after sound onset) regardless of the pupil's overall size at the start of the trial.


a) Participant AB Batch 2 Average PDRs by Trial Type

b) Participant AB Batch 3 Average PDRs by Trial Type

c) Participant AC Batch 1 Average PDRs by Trial Type

d) Participant AC Batch 2 Average PDRs by Trial Type

Figure 4 – Average PDR Traces by Trial Type for Each Batch

Averaged pupil sizes from 1 second before sound onset to 2 seconds after sound onset for habituation trials, oddball1 trials, and oddball2 trials. Blue traces indicate habituation trials (/ka/ sound was played), red traces represent oddball1 trials (/ga/ sound was played; English-native contrast), and green traces represent oddball2 trials (/kha/ was played; non-native contrast in English). Subfigures are a) AB's first batch, b) AB's second batch, c) AC's first batch, d) AC's second batch. Average traces were calculated by averaging pupil size for each sampling point for each trial type for each batch.

The average pupil dilation responses in Figure 4 indicate some important characteristics of the SIPR and the present sample. First, the pupil typically begins to dilate around 250 milliseconds (ms) after sound presentation, a fact that has been demonstrated in previous human SIPR work. Further, as has also been found in previous pupillometry discrimination in humans, the pupil sizes in these average traces seem to increase markedly at 1 second post-stimulus onset.

Finally, a notable characteristic of the average PDR traces in Figure 4 is the less volatile nature of the averaged habituation trial traces for all four batches as compared to the averaged oddball traces. The more extreme peaks and valleys in the averaged traces for oddball trials likely reflect the smaller sample of oddball trials within each batch. However, further work will need to be done on larger datasets to discern whether there may be more variance in oddball trials regardless of sample size.

Having established the visual characteristics of the pupil traces, we then conducted paired t-tests to quantitatively compare pupil sizes before and after sound onset across different trial types. Importantly, the post-onset period is defined as 0.25s-2s after the sound was played, as the pupil dilation response associated with the orienting reflex has a latency of 250ms (Bala et al., 2020). Table 2 reports the mean and standard deviations of the participant's pupil sizes for the periods before and after stimulus onset, partitioned by batch.

A paired-samples t-test was conducted to compare pupil sizes before and after sound onset across all batches. There was a significant difference in the scores for pre-onset (M = 15.89, SD = 5.80) and post-onset (M = 16.79, SD = 4.81) conditions; t(476) = 3.73, p < .001. Table 3 summarizes the results of these t-tests. It shows that the pooled analysis across all batches resulted in a statistically significant difference in pre/post-onset pupil sizes, indicating a likely sound-induced pupillary response. However, the individual batch analyses show variability: AC Batch 1 and AC Batch 2 had significant changes in pupil size before and after sound onset, while AB Batch 2 and AB Batch 3 did not. Therefore, the evidence for a sound-induced pupillary response is stronger when considering all batches together, but less consistent across individual batches. This increased variation across individual batches likely reflects the small sample size and inherent variability in this pilot study.

	Pre-Onset Mean	**Pre-Onset Std. Dev.**	**Post-Onset Mean**	**Post-Onset Std. Dev.**
Participant AB Batch 2	14.81	4.43	15.23	4.08
Participant AB Batch 3	18.51	7.56	18.85	6.74
Participant AC Batch 1	14.02	5.76	15.73	4.79
Participant AC Batch 2	16.20	4.97	17.36	2.77
Pooled Across All Batches	15.89	5.80	16.79	4.81

Table 2 – Pupil Size Mean & Std. Dev. Pre/Post Sound Onset

	Mean Difference between Pre/Post-onset Pupil Size	**Std. Dev. of Difference between Pre/Post-onset Pupil Size**	**t-value**	**Degrees of Freedom (df)**	**p-value**
AB Batch 2	0.42	4.26	1.08	119	.282
AB Batch 3	0.34	7.16	.52	119	.604
AC Batch 1	1.71	5.30	3.54	119	<.001
AC Batch 2	1.16	4.02	3.16	119	.002
Pooled Across All Batches	0.91	5.33	3.73	476	<.001

Table 3 – Paired T-Tests to Discern Sound-Induced Pupil Responses

Preliminary Evidence for Discrimination

While it is possible to use inferential statistics to determine whether participants AB and AC successfully discriminated the oddball stimuli from the habituation stimulus, the reliability of these analyses is not strong given the small sample size. Furthermore, because the pilot study comprises data from two participants in the 10-12-month-old age range, the developmental process of perceptual narrowing is not detectable within this data. However, although statistically significant conclusions are unlikely to result from this small sample, we still analyzed the participants' PDRs to habituation and oddball stimuli to determine whether there are preliminary indications of auditory discrimination.

We calculated each trial's Area Under the Curve (AUC) to investigate the participants' auditory discrimination capability. The AUC indicates the magnitude of the pupil dilation response to habituation and oddball sound presentations. The AUC was calculated for each trial by summing the pupil sizes, starting at stimulus onset. Figure 5 shows the area under the curve for each trial for the four batches included in the pilot study.

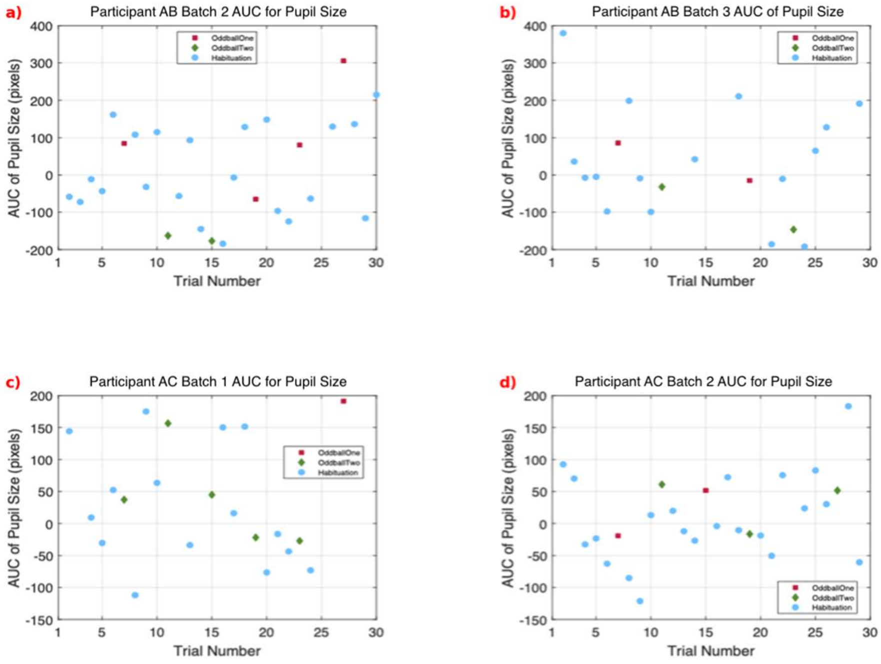


Figure 5 – Area Under the Curve (AUC) for each Trial in All Batches

Figure five shows the area under the curve for each trial from a) AB's first batch, b) AB's second batch, c) AC's first batch, and d) AC's second batch. The area under the curve was calculated by taking the sum of the pupil sizes for each trial starting at sound presentation. Summation started at sound presentation as area under the curve was used to assess differential pupil dilation for habituation vs. oddball sound presentations. Blue dots in the above subplots represent habituation trials, red squares represent oddball 1 (English native contrast) trials, and green diamonds represent (English non-native contrasts).

We conducted a Receiver Operator Characteristic analysis to assess whether the differences in AUC for habituation trials vs. oddball one trial and habituation trials vs. oddball two trials indicate participants discriminating between the given oddball stimulus and the habituation stimulus. The ROC analysis was conducted in MATLAB and adhered the following general process: 1) criterion lines were placed along the range of the AUC data (e.g., for figure 5a, the criterion lines were placed from -200 on the y-axis up to 300) 2) then for each criterion a hit rate was calculated by summing the number of oddball one or oddball two trials with AUC values that fell above the criterion line and dividing it by the total number of oddball one or two trials, 3) finally a false alarm rate was calculated by summing the number of habituation trials with AUCs above the criterion line and dividing that by the total number of habituation trials. Then, for each criterion line, a point is plotted along the ROC curve (see Figure 6) that represents the Hit Rate (also called the Probability of Detection or PODy) divided by the false alarm rate (also called the Probability of false Detection or PODF).

Finally, the overall percent correct (p(c)) was calculated by computing the area under the ROC curve. For two samples drawn from the same population, the expected p(c) is 0.5. Discrimination is indicated when the p(c) equals or exceeds 0.75, or conversely, falls below 0.25. A p(c) below 0.25 suggests that the pupil dilates less in response to a test or oddball trial compared to the habituating sound stimulus. Figures 6b and 6d show a p(c) below 0.25 when comparing PDRs for oddball two to those for the habituation stimulus (an English non-native contrast). Thus, the pupil size difference for PDRs in response to the aspirated oddball two stimulus, relative to the habituation stimulus, reveals statistically significant discrimination. This result appears to contradict our prediction that participants in the 10-12 month age range would no longer discriminate between the oddball two stimulus and the habituation stimulus, as this represents an English non-native contrast. However, as the direction of the difference between two samples does not determine discrimination in ROC analyses, this finding may indicate an attenuated pupil response to the English non-native contrast (oddball two), which aligns with our predicted outcome. Despite this alignment, the result remains unusual, as prior research suggests that the pupil should dilate more to oddball stimuli than to habituation stimuli. For instance, the result could be an artifact of the analysis rather than a novel aspect of the orienting reflex’s tendency to habituate and recover. Therefore, this finding warrants further examination.

While the limited sample size precludes definitive conclusions about participants' auditory discrimination, the ROC analysis yields noteworthy results, particularly for participant AB, desmonstrating the intricate relationship between pupil response variability and auditory stimulus differentiation. Figure 6 visualizes the ROC curves for both batches from participant AB, along with the proportion correct, a metric of how different the PDRs were in response to habituation stimuli vs oddball stimuli. These ROC curves reveal a general trend of greater pupil dilation in response to oddball one compared to the habituation stimulus and reduced pupil dilation in response to oddball two compared to the habituation trials. While the proportion correct for oddball one vs. habituation trials was still below .75, the standard threshold in ROC analyses, it is notable that participant AB showed more robust pupil dilation in response to oddball one than to oddball two. This may be indicative of the fact that this participant is older than ten months and is thus likely attenuating their responsiveness to non-native phonetic contrasts, although a larger sample would be necessary to prove this definitively.

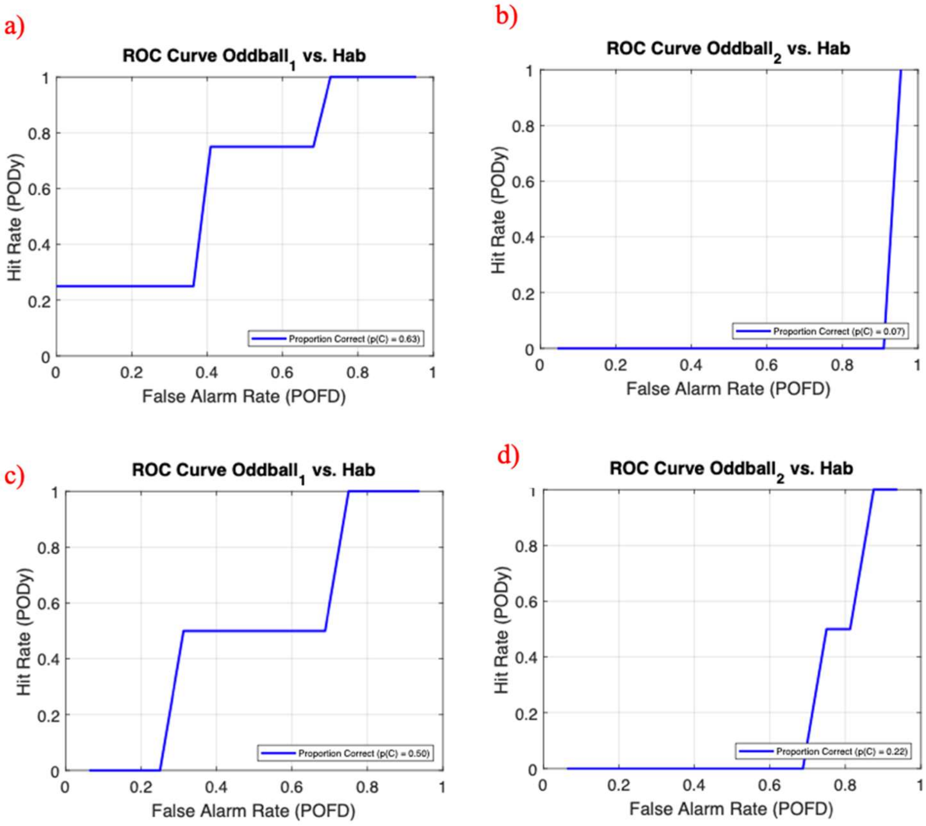


Figure 6 – ROC Curves for Participant AB’s first and second batches:

Figure 6: ROC curves for participant AB’s first and second batches. Subplots: a) ROC curve for batch one, comparing PDRs for habituation trials vs. oddball one trials (English native contrast), b) ROC curve for batch one, comparing PDRs for habituation trials vs. oddball two trials (English non-native contrast), c) ROC curve for batch two, comparing PDRs for habituation trials vs. oddball one trials (English native contrast), d) ROC curve for batch two, comparing PDRs for habituation trials vs. oddball two trials (English non-native contrast). The typically accepted threshold for discrimination is a p(c) at or above 0.75 or equal to or below 0.25.

Discussion

General Discussion

This pilot study was designed to validate the reliability of a pupillometry-adapted oddball paradigm for assessing infant phonetic discrimination. We hypothesized that this method would generally replicate previous studies' findings regarding the developmental course of perceptual narrowing (PN), the process by which infants attenuate their ability to distinguish non-native phonetic contrasts and amplify their responsiveness to native phonetic contrasts (Werker et al., 2012). Based on prior work using the pupillary dilation response (PDR) measure for assessing adult auditory detection and discrimination, we hypothesized that adapting this method for infants may provide a more nuanced view of perceptual narrowing. This is due to the continuous nature of pupillometry data compared to the discrete data from behavioral methods of phonetic discrimination. My role in this project ranged from experiment design, hardware assembly, participant recruitment, running tests to acquire data, writing Python code for data analysis, and conducting a detailed data analysis from two infant participants to conduct a pilot validation study on the phoneme discrimination assay.

While a larger group of 9 infants was tested, I analyzed data from two infants for this thesis. Both these infants were 11 months old, thus likely to be farther along the PN developmental trajectory. Data was collected in two batches of thirty trials for these infants. The batches were designed to test the participants' responses to both English native and non-native phonetic contrasts. It has been previously demonstrated that after ten months of age, infants no longer attend to non-native phonetic contrasts to the same degree as they did a few months earlier (Kuhl, 2004). Therefore, we expected that the data from these participants would indicate attenuated discrimination of phonemes that represent non-native phonetic contrasts and, conversely, would exhibit robust discrimination of phonemes representing native phonetic contrasts. Our findings generally accorded with these predictions, however with the intriguing caveat that infants displayed significant discrimination of non-native contrasts, relative to the habituation stimulus, in the form of a significantly attenuated pupillary response.

Our findings indicate that this adapted oddball paradigm, utilizing PDR data collected during the sound presentation, indeed did elicit responses of differing magnitudes for English native contrasts relative to English non-native contrasts. The small sample of this pilot study precludes the use of inferential statistics. That said, our results clearly suggest, on the one hand, that the two participants demonstrated pupil dilation responses which recovered from their habituated magnitude when they were presented with an oddball phoneme representing an English native phonetic contrast. Also, our participants were in the 10-12 month age range, which is thought to be more progressed through the PN process, and the PDR magnitude displayed noticeably reduced dilation for oddball stimuli representing English non-native contrasts. These findings are congruent with the findings of previous research based on behavioral methods, such as the head-turn procedure. (Kuhl, 2004; Werker et al., 2012).

Thus, the results of this pilot study are promising regarding the utility of this pupillometry-based oddball paradigm for measuring infant phonetic discrimination ability, a critical developmental milestone associated with several other cognitive developmental processes (e.g., attention and executive functioning as indicated by the different PN developmental trajectory in bilingual infants)(Werker & Hensch, 2015). Therefore, this pilot study contributes to the field’s understanding of infant auditory perception, specifically regarding speech, by investigating a novel method that could complement the body of literature on perceptual narrowing that has historically relied on behavioral methods for assessing discrimination. This novel method could help researchers answer new questions about the PN process, such as how the timeline of PN interacts with individual differences, and provide new empirical evidence for how various neurodevelopmental disorders may impact PN. Future research utilizing this novel method with larger sample sizes will be critical for validating this approach and further elucidating this method's utility and capacity for extending the field's understanding of PN beyond what has been uncovered by studies utilizing behavioral approaches.

Limitations

Although this pilot study's findings are encouraging, several limitations must be acknowledged to contextualize the results, allow for appropriate interpretation, and guide future research. Above all, the small sample size of this pilot study limits the degree of confidence with which conclusions can be drawn about the relative sensitivity and robustness of our pupillometry approach to studying phonetic discrimination and perceptual narrowing compared to prior research conducted with behavioral methods. Additionally, since both participants in this pilot study were in the 10-12 month age range, we could not address developmental changes in responsiveness to non-native phonetic contrasts versus native contrasts characteristic of the perceptual narrowing process. This limitation precluded a cross-sectional comparison of younger and older infant responsiveness to non-native contrasts, which should be a key facet of future research utilizing this method to study phonetic discrimination.

Collecting large amounts of data remains challenging, even with this paradigm, which is arguably less cumbersome than behavioral approaches due to the elimination of a training period for infants, required in behavioral methods. The pupillometry method relies on having the infant's eyes in view of the camera, and factors such as blinking or fussiness can limit data quality. Additionally, the first batch from each participant yields higher quality data than later batches, presumably because infants are more attentive and less fussy initially. Given that participant responses to the oddball stimuli habituate with repeated presentation, even when oddball stimuli presentations are spaced out to cope with this habituation, it is likely that in later batches the magnitude of PDRs to oddball stimuli may decrease, as demonstrated in prior research (Bala et al., 2020). Moreover, the presentation of non-native contrasts in earlier batches might undercut attenuation to these contrasts by acting as a learning experience, thereby inadvertently increasing their relevance. Reconciling the diminishing data quality of later batches with the need for comprehensive data collection will be a crucial consideration for future research.

Another limitation of this pilot study is that the small sample size analyzed for this thesis precludes a direct comparison of the relative sensitivity of this measure to behavioral methods. Prior research with adults by Bala and colleagues (2020) has incorporated a receiver operating characteristic (ROC) curve, which effectively converts the continuous data provided by this PDR method to a discrete measure of proportion correct (p(C)). This measure represents the degree to which participants can discriminate the target (the oddball phoneme) from the habituation stimulus. The ROC curve allows for comparing the sensitivity of the PDR-based approach to traditional approaches that yield binary data. However, with the small dataset in this pilot study, ROC analysis is not entirely informative, limiting our ability to make definitive comparisons.

Furthermore, partly due to the proliferation of pupillometry methods in psychological research, there is a lack of consensus regarding how pupillometry data should be analyzed, adding another layer of complexity. Some research uses commercial eye-tracking systems, which obscure how pupil size is calculated. Additionally, custom pupillometry setups are configured differently from one study to another. Consequently, the resulting pupil data can vary widely, even when researchers analyze the same raw dataset of videos (Sirois & Brisson, 2014). This variety of analytical approaches would need to be reconciled if the method outlined in this pilot study is to be utilized widely. Standardizing these analytical methods will be crucial for the validity and reliability of future research using pupillometry to study phonetic discrimination and perceptual narrowing.

Future Directions

The primary directive for future research utilizing this novel approach to studying infant auditory discrimination will be to collect more data. With a larger dataset, many of the limitations of the present pilot study can be effectively mitigated. For instance, with more data, the habituation of PDRs to oddball stimuli can be accounted for by conducting a linear trend analysis to ascertain the degree to which PDR magnitudes in response to habituation and oddball stimuli are diminished with repeated presentation.

Another key aspect for future research is to include participants across both the 6–8-month age range and the 10-12-month age range to allow for cross-sectional analysis. This will help uncover developmental change in both a) attenuation of responses to non-native contrasts and b) the amplification of responses to native contrasts that define the perceptual narrowing process. With this larger and more diverse dataset, future research will be able to utilize an ROC curve analysis to compare the sensitivity of this pupillometry-based approach to prior behavioral approaches.

Once our PDR method has been thoroughly validated with a larger dataset, it will be important to assess its effectiveness in more naturalistic acoustic environments outside of the lab. This will help determine the practical applicability of the method in real-world settings and further validate its utility in studying infant auditory discrimination. Additionally, future research could explore the use of our PDR method with populations thought to be at risk for cognitive developmental delays, such as those with micronutrient deficiencies. This could provide valuable insights into how such conditions impact auditory discrimination and the perceptual narrowing process, potentially leading to early identification and intervention strategies.

Conclusion

This pilot study evaluated the efficacy of a pupillometry-adapted oddball paradigm for assessing infant phonetic discrimination. Our findings indicate that this method elicits distinct pupil responses to English native and non-native phonetic contrasts, aligning with previously documented developmental patterns of perceptual narrowing.

These results highlight the potential of using pupillometry as a non-invasive measure for studying auditory discrimination in infants, contributing to our understanding of early language acquisition and cognitive development. The study of infant language perception is at an exciting juncture, and novel methods like this may help answer new questions about how infants decode their complex acoustic environments. However, limitations such as the small sample size and narrow age range of participants underscore the need for further research with larger and more diverse populations to confirm these findings.

Future research should analyze data from the remaining 7 participants tested in this pilot study, collect more extensive data, and incorporate cross-sectional analysis of different age groups. As the method relies on sounds presented synchronously with a simple animation in tandem with off-the-shelf camera equipment, it is amenable to adaptation for use in the field. For example, sounds can be presented over headphones, and the animation and data acquisition software run on a relatively power tablet computer. A portable test of this kind would be usable for various field studies since it would no longer require a laboratory environment for testing,

Additionally, applying this method to populations at risk for cognitive developmental delays could provide valuable insights and practical applications. The continued development and validation of this pupillometry-based approach hold promise for advancing our understanding of infant auditory perception. This method could become a valuable tool for early diagnosis and intervention, significantly impacting developmental psychology and language acquisition research. By providing a more detailed and nuanced understanding of perceptual narrowing, the PDR approach has the potential to inform educational strategies and therapeutic interventions, ultimately contributing to better developmental outcomes for children.

References

Bala, A. D. S., & Takahashi, T. T. (2000). Pupillary dilation response as an indicator of auditory discrimination in the barn owl. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology*, *186*(5), 425–434. https://doi.org/10.1007/s003590050442

Bala, A. D. S., Whitchurch, E. A., & Takahashi, T. T. (2020). Human Auditory Detection and Discrimination Measured with the Pupil Dilation Response. *Journal of the Association for Research in Otolaryngology*, *21*(1), 43–59. https://doi.org/10.1007/s10162-019-00739-x

Gervain, J., & Mehler, J. (2010). Speech Perception and Language Acquisition in the First Year of Life. *Annual Review of Psychology*, *61*(Volume 61, 2010), 191–218. https://doi.org/10.1146/annurev.psych.093008.100408

Gervain, J., & Werker, J. F. (2008). How Infant Speech Perception Contributes to Language Acquisition: How Infant Speech Perception Contributes to Language Acquisition. *Language and Linguistics Compass*, *2*(6), 1149–1170. https://doi.org/10.1111/j.1749-818X.2008.00089.x

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science*, *143*(3611), 1190–1192. https://doi.org/10.1126/science.143.3611.1190

Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, *154*(3756), 1583–1585.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLoS ONE*, *7*(5), e36399. https://doi.org/10.1371/journal.pone.0036399

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, *69*(1), 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Sirois, S., & Brisson, J. (2014). Pupillometry. *WIREs Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Sokolovc, E. N. (1990). *The orienting response, and future directions of its development*.

Werker, J. F., & Hensch, T. K. (2015). Critical Periods in Speech Perception: New Directions. *Annual Review of Psychology*, *66*(1), 173–196. https://doi.org/10.1146/annurev-psych-010814-015104

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How Do Infants Become Experts at Native-Speech Perception? *Current Directions in Psychological Science*, *21*(4), 221–226.