INFANCY, *12*(2), 119–145

Assessing Speech Discrimination in Individual Infants

Derek M. Houston and David L. Horn
Department of Otolaryngology
Indiana University School of Medicine

Rong Qi
Department of Medicine
Indiana University School of Medicine

Jonathan Y. Ting
Department of Otolaryngology
Indiana University School of Medicine

Sujuan Gao
Department of Medicine
Indiana University School of Medicine

Assessing speech discrimination skills in individual infants from clinical populations (e.g., infants with hearing impairment) has important diagnostic value. However, most infant speech discrimination paradigms have been designed to test group effects rather than individual differences. Other procedures suffer from high attrition rates. In this study, we developed 4 variants of the Visual Habituation Procedure (VHP) and assessed their robustness in detecting individual 9-month-old infants' ability to discriminate highly contrastive nonwords. In each variant, infants were first habituated to audiovisual repetitions of a nonword (*seepug*) before entering the test phase. The test phase in Experiment 1 (extended variant) consisted of 7 old trials (*seepug*) and 7 novel trials (*boodup*) in alternating order. In Experiment 2, we tested 3 novel variants that incorporated methodological features of other behavioral paradigms. For the oddity variant, only 4 novel trials and 10 old trials were used. The stimulus alternation variant was identical to the extended variant except that novel trials were

Send correspondence to Derek M. Houston, 699 West Drive, RR044, Indiana University School of Medicine, Indianapolis, IN 46202. E-mail: dmhousto@indiana.edu

replaced with "alternating" trials—trials that contained repetitions of both the old and novel nonwords. The hybrid variant incorporated elements from both the oddity and the stimulus alternation variants. The hybrid variant proved to be the most successful in detecting statistically significant discrimination in individual infants (8 out of 10), suggesting that both the oddity and the stimulus alternation features contribute to providing a robust methodology for assessing discrimination in individual infants. In Experiment 3, we found that the hybrid variant had good test–retest reliability. Implications of these results for future infant speech perception work with clinical populations are discussed.

Developmental scientists have made significant progress over the last 35 years in delineating speech discrimination capacities in infants as a collective group (Houston, 2005; Jusczyk, 1997). However, there has been much less progress in assessing speech discrimination abilities on an individual basis. Nearly all behavioral paradigms used for infant speech perception are designed to detect general trends in behavior rather than individual differences (Jusczyk, 1997; Werker et al., 1998). In a typical experiment, infants are presented with two sounds and investigators measure responses to determine if infants, on average, behave differently when the speech sound changes compared to when it does not (Werker et al., 1998). Individual infant behavior is often very difficult to predict because of fluctuating internal emotional and physiological states. Also, infants have not acquired the cognitive or linguistic ability to understand what is expected of them during an experiment. Consequently, infant researchers must routinely test a large number of participants for average differences across conditions to emerge from the noise created by these factors.

Although a group design is ideal for discovering general linguistic and cognitive processes in infants, it is not sufficient for clinical research. In many cases, clinical researchers are faced with populations where only a small number of participants can be grouped for analyses or where each test participant presents a unique case. A good example is the population of infants who are deaf and undergo cochlear implantation. These infants may differ with respect to the onset, etiology, and degree of their hearing loss. They may also differ with respect to the age at which they receive a cochlear implantation and the type of linguistic environment (e.g., oral-only or a mix of sign and oral) and habilitation regimen they are immersed in after implantation. Many infants with profound hearing loss are multiply involved, with other conditions that may affect speech perception and language development. Finally, each infant possesses different underlying cognitive and linguistic capabilities. All of these factors contribute to the large differences observed in speech perception and language skills among prelingually deaf children who receive cochlear implants (American Speech-Language-Hearing Association, 2004; Houston, Ying, Pisoni, & Kirk, 2003; Kirk, Miyamoto, Ying, Perdew, & Zuganelis, 2002; Pisoni, 2000; Pisoni, Cleary, Geers, & Tobey, 2000; Sarant, Blamey, Dowell, Clark, & Gibson, 2001; Tobey, Geers, & Brenner, 1994). Developing more precise and reliable behavioral assessment tools for speech discrimination skills in individual infants would provide better means for understanding how these factors contribute to small-group and individual variability.

From a clinical perspective, it is crucial that clinicians have a relatively quick and reliable tool for assessing speech perception and language skills in very young children. This tool would allow clinicians to determine the effectiveness of different interventions and therapy strategies at every appointment and make changes as needed. Currently, clinicians rely on highly subjective parental histories, observation, and behavioral checklists to track the progress of deaf infants with cochlear implants. When these patients are old enough to follow instructions, objective behavioral assessment of speech perception and language skills can be obtained (e.g., Reynell & Huntley, 1985). However, for infants who are too young to follow instructions, there are currently no behavioral methodologies used in standard clinical practice for assessing speech discrimination or any other speech perception skill. There have been some promising advances in developing electrophysiological and neural imaging techniques to measure speech discrimination in infants (Kuhl, 2004). However, further research is needed to cross-validate these techniques with behavioral measures of perceptual discrimination in individual infants before they can be incorporated into viable clinical tools of infant speech discrimination (Purdy et al., 2004).

INFANT SPEECH DISCRIMINATION METHODOLOGIES

Two main types of behavioral methodologies are typically used to investigate infant speech discrimination skills: conditioned head turn (CHT) procedures and visual habituation procedures (VHPs; Werker et al., 1998). The original CHT procedure (referred to as visual reinforcement audiometry) was initially developed to assess infants' auditory thresholds (Moore, Wilson, & Thompson, 1977), and its use was soon extended to also investigate speech discrimination (Eilers, Wilson, & Moore, 1977; Kuhl, 1980). The CHT method relies on conditioning infants to orient to a reinforcer in response to the presence of a change in speech sound (Primus, 1992). After training an infant, the experimenter presents the infant with several test trials to determine if the infant is significantly more likely to orient in response to a change in sound than when there is no change. A major drawback of the CHT is that it often has a high attrition rate, up to 50%, primarily due to many infants not passing the conditioning criterion during the training period (Werker et al., 1998). This may be especially problematic in clinical populations of infants whose cognitive capacities could make conditioning them even more challenging. However, some progress has been made in adapting the CHT for use with clinical populations. Recently, a version of the CHT called the visual reinforcement assessment of the perception of speech pattern contrasts (VRASPAC) has been used successfully in assessing the perception of vowel height in a 17-month-old with a moderate hearing loss (Eisenberg, Martinez, & Boothroyd, 2004).

The other commonly used methodology for investigating speech discrimination is the VHP. The VHP was initially developed for studies of infant visual discrimination (e.g., Cohen, 1969; Kagan & Moss, 1965) and was later modified and used by several investigators for studies of infant auditory discrimination (Best, McRoberts, & Sithole, 1988; Horowitz, 1975; Miller, 1983; Polka & Werker, 1994). The VHP is based on findings that changes in auditory stimuli can affect infants' visual attending behavior to a simple visual display such as a checkerboard pattern (Horowitz, 1975). Thus, to assess discrimination, infants are presented with several trials of a visual display and a repeating sound until their looking time decreases and reaches a habituation criterion. Then, they are presented with a novel sound with the same visual display. An increase in looking time in response to a novel sound is taken as evidence of infants being able to discriminate the two auditory stimuli. The VHP has been used extensively to investigate infants' discrimination of both native and nonnative phonetic contrasts (Best et al., 1988; Polka & Werker, 1994).

An advantage of the VHP when compared to the CHT procedure is that the attrition tends to be somewhat lower (Werker et al., 1998), presumably because habituation is a behavior that does not require conditioning. Because of its low attrition rate, the VHP may be a viable behavioral procedure for investigating speech discrimination in special populations. In a recent study, Houston, Pisoni, Kirk, Ying, and Miyamoto (2003) used the VHP to investigate speech discrimination skills in deaf infants who use cochlear implants. They found that within 6 months after cochlear implantation, deaf infants were able to discriminate discontinuous (*hop hop hop* ...) and continuous (*ahhh*) speech patterns. However, a group design was used comparing the average performance of hearing-impaired infants at various intervals after cochlear implantation with that of normal-hearing infants at various ages. The investigation did not assess speech discrimination skills in individual infants.

LIMITATIONS OF THE VHP

Although the VHP avoids having to condition infants' responses, the VHP methodology allows experimenters to obtain only a very limited number of responses during the test phase. In most cases, infants are presented with one trial of the habituated speech sound and one trial of a novel speech sound (or one trial per novel condition[1]). Infants' average looking times to a checkerboard pattern, paired with

[1]In some experiments, investigators compare different kinds of changes within-subjects. For example, after presentation of audiovisual stimuli, Lewkowicz (2000) presented infants with one novel trial where the visual information changed, one where the auditory information changed, and one where both changed. Thus, there were three novel trials, but each was of a different kind.

either the old or the novel speech sounds, are compared to assess whether infants look longer to the novel speech sound, which would suggest discrimination. With only two test trials, it is difficult to detect a statistically reliable behavioral pattern unless a large group of participants is tested. However, this is often not possible when studying clinical populations.

One possible reason that many VHP experiments are designed with only one novel trial is that the effect of novelty is, by nature, a transitory effect. Indeed, some studies on infant visual habituation have shown that after being habituated with one stimulus and then presented with a novel stimulus, infants demonstrate a recovery of attention to the habituated stimulus (Kaplan & Werner, 1986). Moreover, it has been established that a dishabituating stimulus will itself become a habituating stimulus after repeated exposure (Thompson & Spencer, 1966). Also, because infants can exhibit familiarity preferences under some test conditions (Hunter & Ames, 1989; McMurray & Aslin, 2005), it is possible that after dishabituation, some infants may sometimes switch their preference to the old, familiar stimulus. For these reasons, infants' preference for novelty may not persist over repeated presentations of a novel stimulus. Instead, extending the test phase of a habituation/dishabituation-designed experiment with additional novel and old trials may negate any novelty preference found during the initial two test trials.

PLANS FOR IMPROVING THE VHP

There are potentially enormous research and clinical benefits that could be gained by obtaining sufficient individual data to assess each infant's discrimination skills during a single test session and across test sessions. The investigation was conducted with three goals. The first goal was to determine if infants' preference for a novel stimulus persists after repeated posthabituation presentations of the old and novel stimuli (Experiment 1). The second goal was to develop an infant speech discrimination paradigm that avoids the high attrition rate associated with having a conditioning phase. This paradigm would also need to keep infants on task long enough during the test phase to assess stimulus discrimination in individual infants with statistical reliability within a single session (Experiment 2). The third goal was to assess the test–retest reliability of the new paradigm between sessions (Experiment 3).

Across four variants of the VHP, we used a single perceptual contrast: an audiovisual presentation of a woman producing the nonword *seepug* ([ə]) versus an audiovisual presentation of the same woman producing the nonword *boodup* ([ə]). These two highly contrastive audiovisual stimuli were used to maximize the possibility that all of the 9-month-old infants tested would be able to discriminate the stimuli using both auditory and visual information. Highly contrastive stimuli were used so that each varint of the VHP could be tested for its sensitivity in detecting discrimination in infants. It was assumed that these highly contrastive stimuli would be easily discriminable by infants, using any of the factors by which they varied. Thus, we interpreted each null result as a performance issue rather than a competence issue (i.e., the failure of the procedure to elicit the performance necessary to detect the infants' discrimination capacity).

Each experiment consisted of a habituation phase and a test phase. The habituation phase was identical across experiments. Each infant was presented with repetitions of *seepug* until his or her mean looking time during three consecutive trials was 50% or less than his or her mean looking time during the first three trials. The test phase differed across each experiment. In Experiment 1, infants were presented with seven trials of the old stimulus, *seepug*, and seven trials of the novel stimulus, *boodup*, in alternating order. The results of Experiment 1 served as a baseline for evaluating variants of the VHP. Experiment 2 tested three new variants of the VHP. One integrated an oddity paradigm into the test phase. The second variant borrowed from the Stimulus Alternation Preference Procedure (SAPP; Best & Jones, 1998). A third variant combined elements of oddity and SAPP paradigms to create a hybrid paradigm. To compare these paradigms, 30 infants were randomly assigned so that 10 infants were tested on each variant. This small number of participants resembles a typical scenario for clinical research and presents each variant with a stringent test in which to detect statistically significant effects. By comparing the paradigms under this small *n* condition, we aimed to determine which one would have the best potential for further development and validation for clinical use. Finally, Experiment 3 assessed test–retest reliability by testing a new group of 10 infants on two different days and correlating their performance across test sessions.

EXPERIMENT 1

The aim of the first experiment was to assess the persistence of a novelty effect with highly contrastive audiovisual speech stimuli (*seepug* and *boodup*). If infants consistently looked longer to novel trials across several novel and old trials, then it would be possible to collect enough data from individual infants to reliably assess their ability to discriminate speech contrasts. We refer to this variant of the VHP as the extended variant.

Method

Participants. Ten 9-month-old infants (4 female) were recruited from the greater Indianapolis metropolitan area. All infants passed a newborn hearing screening, had no history of recurrent acute or chronic otitis media, and were not diagnosed with nor suspected of developmental delays by their pediatricians. Other demographic information is presented in Table 1.

TABLE 1
Demographic and Attrition Information for Experiments 1 Through 3

	Age (Months)			Data Not Included (n)				Language Experience (n)		
Experiment/Condition	*M*	*SD*	*Range*	*Crying*	*Fail Fatigue Criteria*	*Fail Habit. Criteria*	*Experimenter or Caregiver Error*	*English Only*	*Primarily English*	*Primarily Other*
Experiment 1: Extended	9.0	.3	8.5–9.6	0	0	0	0	9	0	1
Experiment 2: Oddity	9.0	.3	8.6–9.5	2	1	1	0	10	0	0
Experiment 2: Stimulus alternation	9.0	.3	8.7–9.4	2	3	3	1	6	2	2
Experiment 2: Hybrid	9.1	.2	8.8–9.2	2	3	3	1	9	0	1
Experiment 3: Hybrid	8.7	.2	8.5–9.0	0	0	0	1	9	1	0

Apparatus. The testing was conducted in a custom-designed double-walled Industrial Accounts Company (IAC) sound booth. Infants sat on their caregiver's lap approximately 5 ft in front of a 55-in. wide-aspect TV monitor. The visual stimuli were displayed in the center of the TV monitor at approximately eye level to the infants. The auditory stimuli were presented through both the left and right loudspeakers of the TV monitor. The experimenter observed the infants from a separate room via a hidden, closed-circuit digital camera and controlled the experiment using the Habit software package (Cohen, Atkinson, & Chaput, 2004) operating on a Macintosh G5 desktop computer.

Stimuli. Stimuli were constructed of highly contrastive phonemic and visemic units. Two audiovisual presentations of nonwords were used: *boodup* and *seepug*. Both words conform to the predominant strong–weak stress pattern in English (Cutler & Carter, 1987), making them likely to be heard as possible words by English-learning infants (Jusczyk, Cutler, & Redanz, 1993). Five tokens of *seepug* (four for the habituation phase and one for the test phase) and one token of *boodup* (test phase) were selected from 50 video recordings of a female talker who was instructed to look into the camera and produce the nonwords as if she were speaking to an infant.

The video recordings were edited using FinalCut Pro HD 4.5. Each token was edited so that the face was centered and the sound level was equivalent across tokens (65 ± 5 dB SPL). From each token, a larger QuickTime 6.5 movie file was created, which consisted of 17 repetitions of that token. Within the movie files, each repetition was edited so it appeared to fade in and fade out in to make smooth transitions from one repetition to another.

Four of the *seepug* tokens were used during the habituation phase of the experiment. A fifth token of *seepug* and the *boodup* token were used for the test phase. These two tokens were selected to be equally similar to the four tokens used during the habituation phase in terms of visual characteristics unrelated to the articulation of the nonwords, such as hair placement, eyebrow movement, and so on. The duration of the video portion of each token was 1.83 sec. The durations of the auditory portion of the four *seepug* tokens used during the habituation phase were 1.13 sec, 1.13 sec, 1.27 sec, and 1.27 sec. The auditory portion of the *seepug* token used during the test phase was 1.20 sec, and the auditory portion of the *boodup* token used during the test phase was 0.67 sec.

In addition to these stimuli, two additional stimuli were used in the experiment. A silent video of an infant laughing in the center of the screen was used as an attention getter to orient infants to the TV monitor before each trial. A computer graphic animation, consisting of a geometric shape that moved back and forth accompanied by a sequence of short tones at varying pitches was used to gauge infants' general attention and arousal level before and after the experiment.

Procedure. Infants' looking times to the videos were assessed during the experiment session using the Habit software program (Cohen et al., 2004). The experimenter sat in a control booth located outside of the experimental booth and observed the infants' looking behavior on a TV monitor that had a closed-circuit connection with a hidden video camera in the experiment booth. The experimenter was blind to which stimulus was being presented on each trial and simply pushed a button on the computer keyboard whenever the infants' eyes were oriented toward the video on the TV monitor. Before the onset of each trial, the infants were oriented to the monitor by presenting the attention getter described earlier. Once the infants oriented to the TV monitor, the trial was initiated by the experimenter and continued until the infants looked away from the video for 1 sec or more or until the maximum trial length of 30 sec was reached. The infants' looking time for each trial was calculated as the total amount of time fixated on the visual stimulus during the trial.

Immediately before and after the experiment, infants' general attention and arousal was gauged using a stimulus, described earlier, that was completely unrelated to the experiment stimuli. This is a method used in habituation studies to determine whether decreased looking times in infants could be attributable to habituation to the stimuli or to general fatigue (Cohen & Amsel, 1998; Cohen & Oakes, 1993). To exclude infants who did not maintain their attention during the experiment, one of the inclusion criteria was a looking time during the posttest trial that was at least 50% of the looking time during the pretest trial.

The main experiment consisted of two phases. During the habituation phase, infants were presented with repetitions of four audiovisual *seepug* tokens. Four tokens were used rather than just one with the rationale that any dishabituation that might occur during the test phase would not likely be due to some specific visual change that infants might detect (e.g., orientation of the eyebrows) but due to the auditory aspect of the speech stimuli. Each habituation trial consisted of the attention getter to orient the infants to the center of the TV monitor followed by repetitions of a single token. The trials were ordered so that each of the four tokens occurred once in every set of four trials and no token was repeated across two consecutive trials. The habituation phase continued until the habituation criterion was met, or up to a maximum of 15 trials. The habituation criterion consisted of three consecutive trials in which the mean looking time to the video was 50% or less than the mean looking time during the first three trials.

After the habituation criterion was reached, the test phase began. The experimenter was blinded to when the experiment shifted to the test phase. The test phase consisted of 14 trials: 7 novel trials and 7 old trials. The novel trials consisted of repetitions of a single audiovisual token of the nonword *boodup*. The old trials consisted of repetitions of a novel audiovisual token of the habituation nonword *seepug*. The novel and old trials were presented in alternating order; half of the infants were presented with the novel trial first to control for spontaneous s-posthabituation recovery effects. A schematic of all experimental conditions is displayed in Figure 1.

Results and Discussion

Two sets of analyses were conducted. One set of analyses was used to determine if the infants, on average, demonstrated discrimination of the stimuli throughout the test phase. The second set of analyses was used to determine whether this procedure reliably detected discrimination in individual infants.

Group Data. In the traditional VHP, infants' looking time to a single novel trial is compared to their looking time to a single old trial. Therefore, we first compared looking times to the first novel and the first old test trials in the extended condition. Infants' mean looking times are displayed in Table 2. A repeated-measures analysis of variance (ANOVA) was conducted with trial type (novel trial vs. old trial) as the within-subjects variable and test order (novel first vs. old first) as the

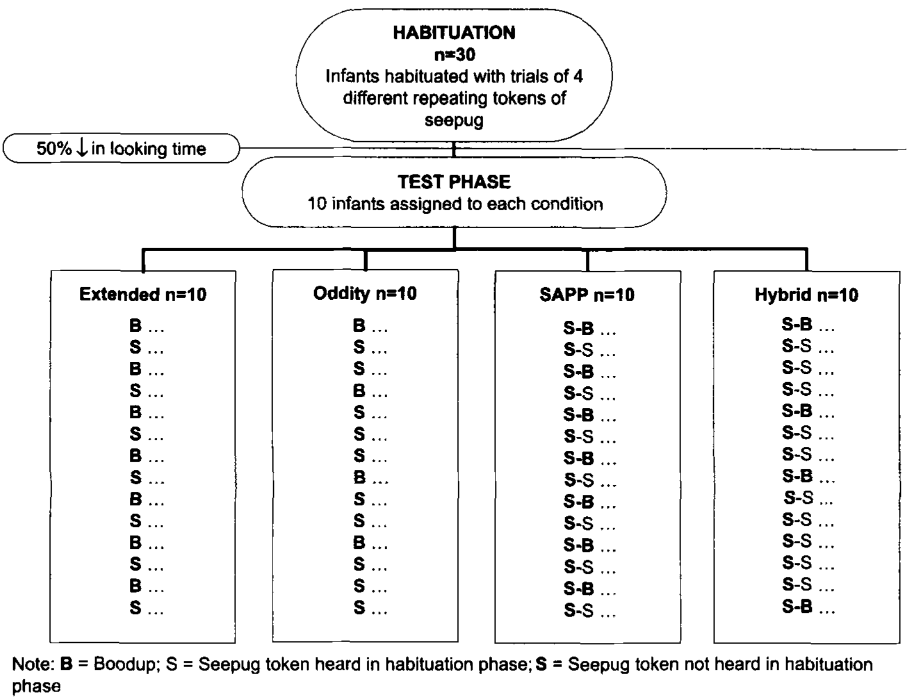

FIGURE 1 A schematic of the experimental conditions. **B** = boodup; S = seepug token presented during habituation phase; **S** = novel seepug token.

TABLE 2
Mean Number of Habituation Trials and Mean Looking Times During the Habituation and Test Phases of Experiments 1 Through 3

	Habituation Rate		*All Habituation Trials*		*Last Habituation Trial*		*First Novel Trial*		*First Old Trial*		*All Novel Trial*		*All Old Trial*	
Experiment/Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experiment 1: Extended	8.1	2.8	14.0	5.9	6.5	2.9	17.0	9.4	11.3	8.0	9.7	5.1	7.9	3.4
Experiment 2: Oddity	8.5	3.1	17.9	6.8	6.3	4.8	19.2	11.4	11.9	8.7	16.8	8.3	9.9	3.5
Experiment 2: Stimulus alternation	6.4	1.8	15.8	3.7	7.1	2.9	22.9	9.8	17.5	11.6	16.3	7.4	11.4	4.2
Experiment 2: Hybrid	7.0	2.7	15.4	5.1	6.0	3.2	21.4	10.2	7.6	5.3	20.4	7.9	7.9	2.3
Experiment 3: Day 1	9.3	3.3	15.5	3.5	7.9	4.0	18.5	9.5	7.3	1.5	17.7	7.5	7.3	1.5
Experiment 3: Day 2	6.4	2.8	10.7	4.5	5.2	4.0	17.0	10.4	6.8	4.2	14.2	6.4	6.0	3.0

Note. Values for habituation rate represent number of trials to reach habituation criterion. All other values represent seconds.

between-subject variable. There was no significant main effect of trial type, $F(1, 8) = 2.80$, $p = .13$, $\eta^2 = 0.26$, and no significant Trial Type × Test Order interaction ($F < 1$), suggesting that this test was not sensitive enough to detect discrimination of *seepug* and *boodup* in a small group of 9-month-olds when only the first two test trials were analyzed. Although it is surprising that infants did not look significantly longer to the novel trial, this result is advantageous for comparing discrimination results across conditions because it indicates that we did not encounter a ceiling effect.

Next, to determine how increasing the number of test trials influences the results in a VHP, we compared infants' mean looking times to the seven novel trials and to the seven old trials (see Table 2). A paired t test indicated that this difference also did not reach statistical significance, $t(9) = 1.68$, $p = .13$, Cohen's $d = 0.53$, suggesting that additional test trials did not make this version of the VHP more sensitive at detecting discrimination in a small number of 9-month-olds.

To more closely examine the effect of adding trials on infants' novelty effect, a mixed model with trial type (novel vs. old) as a fixed factor, trial number (1–14) as a continuous variable, and participant as a random factor was used to analyze the data. This time, the main effect of trial type, $F(1, 136) = 5.52$, $p = .02$, $\eta^2 = 0.03$, was small but statistically significant, suggesting that after factoring in the effect of trial number and thus increasing the degrees of freedom, infants' preference for the novel trials did reach statistical significance. There was also a main effect of trial number, $F(1, 136) = 27.54$, $p < .0001$, $\eta^2 = 0.16$, reflecting a trend of shorter looking times across the trials. The Trial Type × Trial Number interaction did not reach statistical significance, $F(1, 136) = 1.34$, $p = .25$, $\eta^2 = 0.01$, suggesting that infants' novelty preference did not change significantly across trials.

Individual Data. To determine whether adding 12 additional novel and old trials allowed us to detect discrimination of stimuli in infants with statistical reliability, we subjected each infant's looking times to the seven novel trials and to the seven old trials to an autoregression analysis. An autoregression model is a special type of analysis of covariance model used with time series data where the dependent variables could incur correlation in the order in which data are collected. We used the first-order autoregressive model in which the current looking time of a trial is modeled to be dependent on both the specific condition of the trial (novel or old) and the looking time of the previous trial. Parameter estimates and hypothesis testing are conducted using the maximum likelihood approach. An autoregressive model takes into account potential correlations in looking times between trials from the same infant (Chatfield, 2003). Table 3 summarizes the looking times and the results of the autoregression analyses for each participant. The looking times are followed by the p value of the stimulus comparison (novel vs. old) after adjusting for the autoregressive correlation. Only 1 infant demonstrated a looking time preference for the novel trials that was statistically significant. The looking time

TABLE 3

Individual Participant Raw Looking Times and Data Analyses for Experiments 1 and 2

Extended (Experiment 1)				*Oddity (Experiment 2)*				*Stimulus Alternation (Experiment 2)*				*Hybrid (Experiment 2)*			
Participant	*Novel*	*Old*	*p_adj*	*Participant*	*Novel*	*Old*	*p_adj*	*Participant*	*Novel*	*Old*	*p_adj*	*Participant*	*Novel*	*Old*	*p_adj*
1032	**4.67**	**8.83**	**.01**	682	8.04	9.44	.63	824	3.76	7.97	.17	1924	6.15	6.96	.86
1054	6.63	7.20	.28	668	12.43	13.06	.34	1443	14.59	15.31	.76	**822**	**10.03**	**4.17**	**.0003**
1202	2.36	2.40	.93	**889**	**5.18**	**4.23**	**.03**	816	19.01	18.24	.66	772	16.65	9.5	.14
1134	10.21	7.80	.36	903	10.45	8.56	.53	762	8.67	6.74	.53	**1402**	**15.85**	**6.01**	**.002**
999	8.74	6.21	.20	1033	13.68	10.57	.58	1166	16.23	12.67	.37	**837**	**23.02**	**10.96**	**.002**
1225	*8.40*	*5.74*	*.07*	*834*	*16.18*	*7.62*	*.07*	836	10.54	5.76	.29	**1300**	**26.7**	**11**	**.004**
1118	13.37	9.97	.31	**1014**	**29.80**	**17.34**	**.03**	1069	16.66	10.51	.13	**1184**	**22.8**	**6.4**	**<.0001**
929	*11.61*	*7.84*	*.06*	**773**	**22.20**	**8.39**	**.04**	1004	20.89	11.36	.14	**768**	**25.8**	**8.38**	**<.0001**
1008	9.90	5.89	.34	**888**	**24.65**	**9.54**	**.01**	**863**	**28.56**	**16.34**	**.003**	**900**	**29.63**	**9.16**	**.002**
1186	**20.87**	**12.71**	**.03**	**931**	**25.63**	**10.47**	**.003**	**771**	**23.81**	**9.13**	**<.0001**	**789**	**27.35**	**6.41**	**<.0001**

Note. The *p*_adj column contains the *p* values of the stimulus condition (novel vs. old) after adjusting for the autogressive parameter value. Those reaching significance criterion are bolded and those approaching are italicized.

difference of 2 other infants approached significance, and 1 infant showed a significant preference for the old trials. These results suggest the extended variant of the VHP does not provide a more sensitive measure for discrimination in infants than a traditional VHP, and it does not detect discrimination in individual infants.

EXPERIMENT 2

The previous experiment showed it is possible to demonstrate statistically reliable speech discrimination in a small proportion of infants. However, the looking patterns of most infants were not consistent enough to show statistically significant discrimination, even though all infants could presumably discriminate the stimuli. The goal of the second experiment was to modify the VHP so the effect of novelty would be more robust and persist over the entire test phase.

The test phase of the VHP was modified in three ways, and each infant was randomly assigned to one of three test conditions. In the oddity condition, we presented infants with the novel stimulus on 4 trials and the old stimulus on 10 trials in pseudo-random order. The logic behind this variant is that if a stimulus is infrequent, it should be relatively more salient to infants than a stimulus that is more frequent. The idea that individuals may respond differently to infrequent and frequent stimuli has been heavily exploited in sensation and perception studies using electrophysiological techniques in which neural responses to a less frequent (oddball) stimulus are compared to responses to a more frequent (standard) stimulus (see Picton, Alain, Otten, Ritter, & Achim, 2000, for a review). Only a handful of studies have used this technique with behavioral paradigms (Bountress, Sever, & Williams, 1989; Doehring, 1969; Strange, Polka, & Dittmann, 1986), and none that we know of with infants. However, it is possible that the physiological response to infrequency may also manifest itself as a behavioral response, such as increased looking time.

The stimulus alternation condition presented infants with a modified version of the SAPP developed by Best and Jones (1998). SAPP capitalizes on the basic finding that infants show more attention (i.e., habituate more slowly) to more complex stimuli than to simple stimuli (Cohen, DeLoache, & Rissman, 1975; Richard, Normandeau, Brun, & Maillet, 2004; Slater, Rose, & Morison, 1984). Best and Jones (1998) found that infants exhibit longer looking times when presented with alternating speech sounds (e.g., [ba pa ba pa …]) than when presented with repetitions of the same speech sound (e.g., [ba ba ba …]). To compare the stimulus alternation condition to the extended condition, we modified Best and Jones's SAPP so that the habituation phase and number of test trials were the same as those in the variants used in the previous two experiments. Like the extended condition, the test phase consisted of seven old trials and seven novel trials. However, in the stimulus alternation condition *seepug* and *boodup* alternated on novel trials.

The hybrid condition was a combination of the oddity and stimulus alternation conditions. Infants were presented with 4 alternating trials and 10 old trials in pseudo-random order.

Method

Participants. Thirty 9-month-old infants were recruited from the greater Indianapolis metropolitan area. Ten infants were randomly assigned to each of three conditions: the oddity condition (7 female), the stimulus alternation condition (4 female), and the hybrid condition (6 female). All infants had passed a newborn hearing screening, had no history of recurrent acute or chronic otitis media, and were not diagnosed with nor suspected of developmental delays by their pediatricians. Additional demographic and attrition rate information is presented in Table 1.

Stimuli and Apparatus. For the oddity condition, the stimuli and apparatus were identical to the stimuli and apparatus in Experiment 1 for both the habituation and test phases. For the stimulus alternation and hybrid conditions, the stimuli used during the habituation phase were also identical to the previous experiment. However, the stimuli used during the test phase were different. For the old-alternating trials, two audiovisual tokens of *seepug* were presented in alternating order; one was the same token used during the test phase of Experiment 1 (*seepug5*), and the other was one of the four tokens from the habituation phase. This was counterbalanced across infants. For the novel-alternating trials, *boodup* and *seepug5* were presented in alternating order. For both the old-alternating and novel-alternating trials, *seepug5* was always presented first. The interstimulus intervals (ISIs) of these alternating stimuli were identical to the ISIs of the repeating tokens in the other conditions. The reason for having two alternating tokens for both the old-alternating and novel-alternating trials was to ensure that a looking time preference for the novel-alternating trial would be due to detection of the type variation and not simply a token variation.

Procedure. A schematic of each condition is presented in Figure 1. The habituation phase in all three conditions was identical to Experiment 1. The first two trials of each test phase were comprised of one old trial and one novel (oddity) or alternating (stimulus alternation, hybrid) trial. The order of trial types was counterbalanced across infants. For the oddity condition, the last 12 trials were comprised of three novel trials and nine old trials in pseudo-random order with the caveat that novel trials did not occur consecutively. The pseudo-random ordering differed for each infant. For the stimulus alternation condition, the last 12 test trials were the same as in Experiment 1 except that novel and old trials were replaced with novel-alternating and old-alternating trials. For the hybrid condition, the trial ordering was the same as in the oddity condition but the stimuli were the same as in the stimulus alternation condition.

Results

Group Data. Analyses were conducted to test whether any of the new variants were more sensitive to infants' discrimination abilities than the extended variant. To evaluate whether there was any sampling bias among groups of infants assigned to Experiment 1 and the three conditions of Experiment 2, the habituation rate and mean looking times during the habituation phase were compared across groups of infants because the habituation phase was identical for all infants. Infants' habituation rates (i.e., number of trials to reach the habituation criterion) and mean looking times during the habituation phase are displayed in Table 2. An ANOVA revealed that these habituation rates did not differ significantly across groups of infants, $F(3, 40) = 1.37, p = .27, \eta^2 = 0.10$. The mean looking times during the habituation trials also are displayed in Table 2. An ANOVA revealed that the looking times during the habituation phase did not differ significantly across groups of infants, $F(3, 40) < 1$. The results from these analyses suggest that baseline attentional and looking behaviors do not differ significantly across groups of infants. Thus, any differences in performance during the test phase across groups of infants are more likely to be due to differences in testing conditions rather than a selection bias between groups.

First we compared infants' looking times to the first novel and first old trials of the test phase. Looking times are displayed in Table 2. A repeated-measures ANOVA was conducted with trial type (novel trial vs. old trial) as the within-subjects variable and test order (novel first vs. old first) and condition (extended, oddity, stimulus alternation, and hybrid) as the between-subject variables. There was a main effect of trial type, $F(1, 32) = 21.75, p < .001, \eta^2 = 0.36$, and no other significant main effects or interactions. The results suggest that when a large number of participants is tested, these VHP variants are able to detect discrimination in infants with only two trials and that there are no significant differences in discrimination sensitivity across conditions.

Next, we compared infants' mean looking times to all novel and old trials across conditions. Because there were different numbers of novel and old trials across conditions, a mixed model with trial type and condition as fixed factors, trial number as a continuous variable, and participant as a random factor was used to analyze the data. Infants' mean looking times are presented in Table 2. There were main effects of trial type, $F(1, 548) = 126.71, p < .0001, \eta^2 = 0.16$, trial number, $F(1, 548) = 38.69, p < .0001, \eta^2 = 0.05$, and condition, $F(3, 548) = 4.59, p = .004$, $\eta^2 = 0.02$. There was a significant Condition × Trial Number interaction, $F(3, 548) = 8.10, p < .0001$, s used to analyze the $\eta^2 = 0.03$, which reflects a difference across conditions in the amount of looking time decrease across trials. Inspection of the data in Table 2 suggests that this interaction may be due to infants maintaining looking times more in the oddity and hybrid conditions than in the stimulus alternation and extended conditions. There was also a significant Trial Type × Condition interaction, $F(3, 548) = 14.11, p < .0001, \eta^2 = 0.05$, suggesting that when all of the trials were analyzed there was a difference in novelty preference across conditions.

To more closely examine the difference in novelty preference across conditions and to determine if one of the new VHP variants detects discrimination in infants more robustly than the extended variant, additional mixed model analyses were conducted comparing each of the new variants to the extended variant. For each comparison, the main effects of trial type and trial number persisted. In addition, the Trial Type × Condition interaction was significant for the extended versus hybrid condition comparison, $F(1, 272) = 13.30, p = .0003, \eta^2 = 0.04$, and not for the extended versus oddity condition or the extended versus stimulus alternation condition comparisons (Fs < 1). Further analyses revealed that the novelty preference was significantly greater in the hybrid condition than in the oddity condition, $F(1, 272) = 4.01, p = .046, \eta^2 = 0.01$, or the stimulus alternation condition, $F(1, 272) = 5.47, p = .02, \eta^2 = 0.02$, and that the novelty preference did not differ significantly between the oddity and stimulus alternation condition ($F < 1$). These results suggest that infants exhibited more robust looking time differences in the hybrid variant of the VHP than in any of the other variants and that neither the oddity nor the stimulus alternation variant resulted in better performance than the extended variant.

Individual Data. To determine whether individual infants showed statistically reliable looking time differences, we subjected each infant's looking times to an autoregression analysis as in Experiment 1. The looking times and autoregression p values are displayed in Table 3 with infants showing a statistically significant discrimination listed in bold. There was a significant effect of condition on discrimination rate among the three conditions ($p = .006$, Fisher's Exact test). Comparing the discrimination rate of the new variants to the extended variant, infants in the hybrid condition were significantly more likely to show discrimination than infants in the extended condition ($p = .006$). In contrast, infants in the oddity condition ($p = .14$) and infants in the stimulus alternation condition ($p = 1.0$) were not significantly more likely to show discrimination than infants in the extended condition.

There were several infants who completed testing but were not included in the analyses due to failure to meet the habituation or fatigue criteria. To examine whether these criteria are necessary for demonstrating individual discrimination, we submitted the looking times of the excluded infants to autoregression analyses. The 2 infants in the oddity condition who were not included showed statistically significant discrimination. Of the 6 infants in the stimulus alternation condition who were not included, 3 showed statistically significant discrimination. Of the 6 infants in the hybrid condition who were not included, 5 showed statistically significant discrimination. These results show that the infants excluded from our analyses exhibited very similar performance as the infants included in our analyses, suggesting that effects during the test phase are robust to habituation failure and fatigue.

Discussion

The findings suggest that combining oddity and stimulus alternation paradigms results in a more robust infant speech discrimination procedure than a procedure that simply presents additional novel and old trials. Moreover, the hybrid variant was able to detect statistically significant discrimination in individual infants much more often than the extended variant. Thus, the hybrid variant of the VHP appears to hold promise as a useful clinical and research tool for detecting discrimination in individual infants. But will the hybrid variant also be able to be used clinically to track infants' discrimination performance over time? To address this question, it is necessary to assess the test–retest reliability of the hybrid variant.

EXPERIMENT 3

In this experiment, we examined the test–retest reliability of the hybrid variant by comparing infants' performance across two testing sessions separated by 1 to 3 days.

Method

Participants. Ten 9-month-old infants (5 female) were recruited from the greater Indianapolis metropolitan area. All infants had passed a newborn hearing screen, had no history of recurrent acute or chronic otitis media, and were not diagnosed with or suspected of developmental delays by their pediatricians. Additional demographic and attrition information are presented in Table 1.

Stimuli. The stimuli were identical to those used in the hybrid condition of Experiment 2.

Procedure. Both the habituation and test phases were identical to the hybrid condition described in Experiment 2. Infants were tested using the same procedure on two separate days. The first two trials of each test phase were comprised of one old trial and one novel alternating trial. Order of the first two trials was counterbalanced across infants and across days.

Results

Group Data. Analyses were performed to replicate the findings of Experiment 2 as well as to determine test–retest reliability of this paradigm. Looking times during the habituation and test phases are displayed in Table 2.

To evaluate the effect of repeat testing on habituation, we subjected the habituation rate data and the mean looking time data from both days of testing to two separate repeated-measures ANOVAs with test day (1, 2) as the repeated-measures variable. The analyses revealed a significant difference in mean looking time, $F(1, 9) = 15.23$, $p = .004$, $\eta^2 = 0.63$, and a habituation rate difference that approached significance, $F(1, 9) = 4.09$, $p = .07$, $\eta^2 = 0.31$, suggesting that infants had shorter looking times and habituated faster during the habituation phase on the second day than on the first day of testing.

An ANOVA of infants' looking times to the first two test trials across both days revealed a significant main effect of trial type, $F(1, 9) = 27.13$, $p < .001$, $\eta^2 = 0.55$, but no main effect of test day or Trial Type × Test Day interaction (Fs < 1). Similarly, when infants' looking times to all of the novel and old trials were subjected to the same analyses, there was a significant main effect of trial type, $F(1, 9) = 34.28$, $p < .001$, $\eta^2 = 0.61$, whereas the main effect of test day did not reach significance, $F(1, 9) = 3.02$, $p = .12$, $\eta^2 = 0.04$, nor did the Trial Type × Test Day interaction, $F(1, 9) = 1.62$, $p = .24$, $\eta^2 = 0.01$. The results of these analyses suggest that, as a group, infants performed similarly across both test sessions and that they robustly demonstrated discrimination on both days.

Individual Data. Although similar group results across test sessions suggests some degree of reliability of the hybrid variant, they do not reveal the degree to which individual infants performed similarly across testing sessions. As can be seen in Table 4, 8 of 10 infants showed statistically significant discrimination on each of the 2 days of testing. When combining the data from both sessions, 8 out of 10 infants showed statistically significant discrimination and an additional infant's performance approached statistical significance. To test for correlations of individual infant's preferences across test sessions, we created a variable that would take into account the infant's baseline attention during each test session. For each day of testing, a preference quotient was calculated as the difference in mean looking times to novel and old trials divided by the mean looking time to all 14 trials. A Pearson's correlation was then performed using the preference quotient for Day 1 and Day 2 of testing (see Figure 2). We found a significant, moderately strong correlation between the preference quotients on Day 1 and Day 2 of testing ($r = .653$, $p = .041$), suggesting that infants'

TABLE 4
Individual Participant Raw Looking Times and Data Analyses for Experiment 3

	Day 1				*Day 2*				*Both Days*			
Participant	*Novel*	*Old*	*p_adj*	*PQ*	*Novel*	*Old*	*p_adj*	*PQ*	*Novel*	*Old*	*p_adj*	*PQ*
2005	8.90	7.23	.80	0.22	6.63	5.73	.81	0.15	7.77	6.48	.47	0.19
2629	**16.00**	**6.50**	**.003**	**1.03**	13.25	6.65	.25	0.77	**14.63**	**6.58**	**.0077**	**0.91**
2617	**17.75**	**8.80**	**.009**	**0.79**	**22.88**	**8.96**	**.01**	**1.08**	**20.32**	**8.88**	**.0033**	**0.94**
2017	**16.65**	**6.16**	**.02**	**1.15**	**18.28**	**4.11**	**.0008**	**1.74**	**17.47**	**5.14**	**<.0001**	**1.42**
2054	7.18	4.32	.33	0.56	**3.83**	**1.91**	**<.0001**	**0.78**	*5.51*	*3.12*	*.09*	*0.63*
2665	**29.30**	**7.64**	**<.0001**	**1.57**	**19.43**	**6.85**	**.006**	**1.20**	**24.37**	**7.25**	**<.0001**	**1.41**
2066	**12.73**	**7.36**	**.02**	**0.60**	**10.18**	**4.96**	**.01**	**0.81**	**11.46**	**6.16**	**.0008**	**0.69**
2644	**29.83**	**9.62**	**.0001**	**1.31**	**10.13**	**3.18**	**<.0001**	**1.34**	**19.98**	**6.4**	**<.0001**	**1.32**
2638	**20.15**	**7.52**	**.0003**	**1.13**	**20.60**	**12.22**	**.03**	**0.57**	**20.38**	**9.87**	**.0001**	**0.82**
2658	**18.83**	**7.97**	**.02**	**0.98**	**16.45**	**5.16**	**.005**	**1.35**	**17.64**	**6.57**	**.0004**	**1.14**

Note. PQ = preference quotient. Values for novel and old represent seconds. Adjusted *p* values that reached significance criterion are bolded and those that approached significance are italicized.

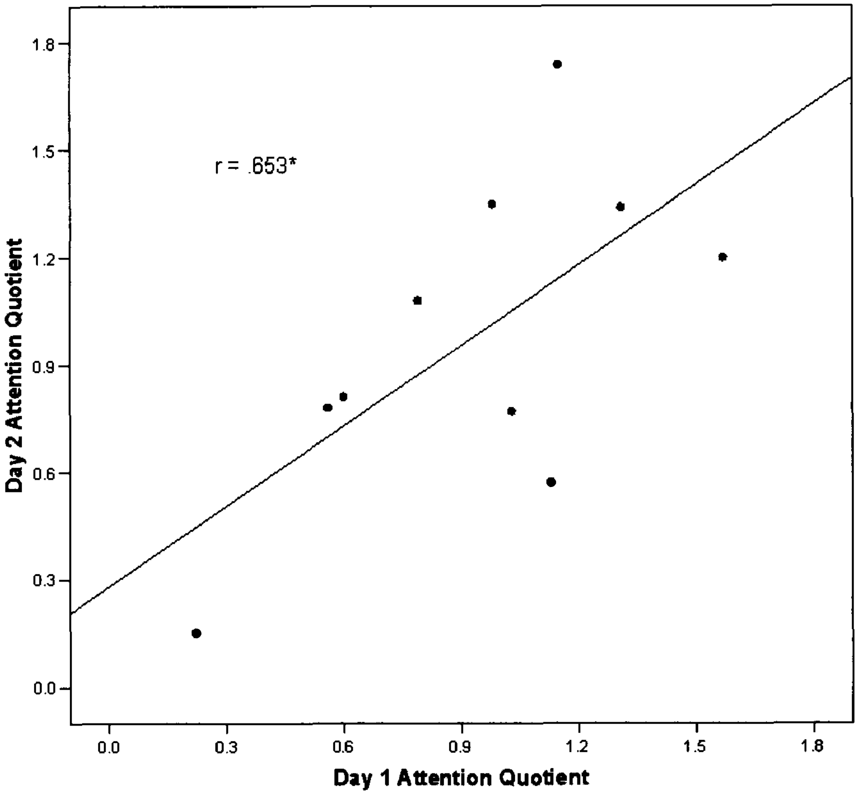

FIGURE 2 A scatterplot and regression line of infants' preference quotient on Day 1 and Day 2 of Experiment .

performance on Day 1 significantly predicted their performance on Day 2 in the hybrid variant.

Discussion

These findings suggest that the hybrid variant provides reliably robust discrimination in infants across test sessions. The consistent performance of infants across testing sessions that were 1 to 3 days apart suggests that the hybrid variant may be a clinically useful tool to assess the improvement in the speech perception abilities of individual infants over time, at least in the context of this easily discriminable contrast.

GENERAL DISCUSSION

The goal of the experiments reported in this article was to test variants of the classic VHP that might provide a means for reliably detecting discrimination in individual infants. The need for such a methodology is considerable. Clinicians who work with populations of infants that have difficulty acquiring spoken language need ways of evaluating whether or not the therapy strategies they are using are effective. A methodology for assessing speech discrimination in a clinical setting must have a low attrition rate and be quick to implement. The variants of the VHP that we employed took about 5 to 8 min for each testing session.

The attrition rate, averaged across the experiments, was about 28%, which is average for infant speech perception paradigms (Werker et al., 1998). However, most of the 22 infants not included in the analyses were removed because of failure to reach the habituation criterion ($n = 7$) or failure to pass the fatigue criterion ($n = 7$). Although these infants were not included in the original analyses, they did complete testing and we analyzed their data separately. In the oddity variant, 2 out of 2 excluded infants showed statistically significant discrimination; in the stimulus alternation variant, 3 out of 6 excluded infants showed discrimination; and in the hybrid variant, 5 out of 6 excluded infants showed reliable discrimination. The discrimination rates of the excluded infants are about the same as for the infants included in the analyses. Thus, passing the fatigue and habituation criteria does not seem to be critical for demonstrating discrimination in these paradigms. The critical element for displaying discrimination appears to be completing the experiment without crying, which over 92% of the infants did.

The hybrid variant, which incorporated elements of the oddity and stimulus alternation variants, elicited greater mean looking time differences than all of the other variants. Neither the oddity element nor the stimulus alternation element alone elicited significantly greater looking time differences than the extended variant. This pattern of results suggests that both the oddity and stimulus alternation elements contributed to infants' superior discrimination performance in the hybrid variant than the extended variant. Given that performance did not differ significantly between oddity and stimulus alternation variants, both elements may have contributed equally to the superior performance of the hybrid variant. However, it is possible that with more participants in each condition, differences between the oddity and stimulus alternation variants would be found.

It is possible that performance differences between the hybrid variant and other variants reflected a sampling bias rather than actual differences in speech discrimination sensitivity. Indeed, infants in the stimulus alternation variant looked much longer, on average, to the first old trial than infants in the hybrid variant even though the conditions of the variants were identical up to that point. However, the similar mean looking times during the habituation phase and the similar habituation rates do not suggest group differences between these infants. An important difference between the habituation data and the first old trial data is that the former were obtained from an average of 6.7 trials whereas the latter were obtained from a single trial. The difference in looking times to the first old trial across the hybrid and the stimulus alternation variants is likely to be spurious and highlights the potential problems of interpretation when small groups of infants are tested with very few test trials. It also emphasizes the need to develop new methods for obtaining multiple data points from each participant in situations where testing large numbers of participants is not feasible.

The pattern of looking times across the four variants suggests that several interacting factors play a role in infants' attention and preferences. Infants' overall mean looking times during the test phases were shorter in the extended variant than in the other two variants. This suggests that introducing elements of complexity to the task, either by stimulus alternation or an oddity presentation, elicited longer looking times from the infants. Stimulus alternation capitalizes on infants' spontaneous preference for more complex stimuli (Cohen et al., 1975; Richard et al., 2004; Slater et al., 1984), and it likely enhances the effect of novelty, as evidenced by significantly longer looking time differences in the hybrid variant than in the oddity variant. The oddity element likely contributes to infants' novelty by maintaining the effect of novelty as evidenced by the maintenance of looking times to the novel trials in the hybrid and oddity variants compared to the stimulus alternation variant.

The differences in looking times to the old trials across variants suggest that infants' looking times to one stimulus are not necessarily independent of their looking times to another stimulus. Three interacting factors may have played a role in infants' looking times during the old trials. First, according to the dual-process theory of habituation (Kaplan & Werner, 1986), a novel stimulus elicits a general arousal, or sensitization, that can result in an increase in infants' attention to a habituated stimulus. Thus, infants' looking times to the old stimuli in these variants of the VHP are likely to be affected by the intensity of their responses to the novel stimuli. Second, the number of old trials during the test phase likely affected mean looking times, because more old trials should result in more habituation to the old stimulus. Finally, the degree to which they preferred the novel stimulus may have influenced their attention to the unpreferred stimulus. Further work will be necessary to understand how these factors interact, as has been done recently in studies of infant color vision (Civan, Teller, & Palmer, 2005).

The robust discrimination performance in the hybrid variant was replicated in Experiment 3. Moreover, we found a significant positive correlation in infants' performance across two different testing sessions that were 1 to 3 days apart. Given that this contrast is likely to be easy enough for all infants to discriminate, it might be expected that the variance in performance would be low. Thus, the moderately strong correlation reveals an impressive consistency of performance in infants. More important, 8 out of 10 infants showed stable performance across testing sessions: 7 showed statistically significant discrimination on both days and 1 did not show discrimination on both days. This degree of reliability would provide clinicians with a moderate amount of confidence in the results of a speech discrimination test implemented once and, potentially, a high amount of confidence when results are taken from multiple visits. However, further work is needed to assess the reliability of the hybrid variant when implemented more than twice. The findings of Experiment 3 also provide additional counterevidence to the possibility that the superior performance in the hybrid variant was simply due to a sampling bias.

Very few studies have examined the test–retest reliability of speech discrimination measures in individual infants or have tracked infants' speech discrimination skills longitudinally (but see Nozza, Miller, Rossman, & Bond, 1991; Werker & Tees, 1984). This is probably because research has focused on understanding the general development of speech perception skills in infants and not as much on individual differences or within-subject developmental change. Thus, the reliability of individual performance has been of little concern. Given the many extraexperimental factors that may affect infants' performance during a given test session, it would be reasonable to assume that test–retest reliability of individual infants is low and that variability in performance may simply reflect noise in the data. However, there has recently been an increased interest in individual variability on speech perception measures (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Newman, Bernstein Ratner, Jusczyk, Jusczyk, & Dow, 2006; Tsao, Liu, & Kuhl, 2004). For example, Kuhl, Tsao, and colleagues have reported correlations between infants' performance on speech discrimination tasks and their later receptive and expressive vocabularies (Kuhl et al., 2005; Tsao et al., 2004). Newman et al. (2006) recently found that infants' performance on tasks related to segmenting words from fluent speech predicted their expressive vocabularies at 2 years of age and their general linguistic skills at 4 to 6 years of age.

These recent findings represent an exciting new direction of infant speech perception research. Unexplained variance in performance among infants may turn out to be meaningful for predicting later language outcomes, which could have important clinical implications for early detection and intervention of language difficulties. Developing more robust and reliable techniques for assessing individual infants' speech perception performance would contribute significantly to identifying early predictors of language disorders in addition to providing clinicians with new methods for assessing the progress of their patients in the clinic. The findings with the hybrid variant of the VHP represent a promising step toward those goals. However, additional reliability and validity testing is necessary to further assess the viability of the hybrid variant of the VHP as a clinical tool for assessing infant speech discrimination.

ACKNOWLEDGMENTS

This research was supported by a National Institute for Deafness and Other Communication Disorders [NIDCD] Research Grant (R01 DC006425) to Derek M. Houston and by an NIDCD Training Grant (T32 DC00012) to David Pisoni. We thank Carrie Hansel, Kabreea York, Katherine Colden, Rebecca Benjamin, and for assistance with data collection and manuscript proofreading; we thank David Pisoni, Tonya Bergeson, Rachael Holt, Roberta Golinkoff, Karen Kirk, and Elizabeth Ying for helpful discussions and critiques of this work; we thank Kasi McCune for recording the stimuli; and we thank the other members and participants (infants and their parents) of the Infant Language Laboratory at the Indiana University School of Medicine for their participation in this research.

REFERENCES

American Speech-Language-Hearing Association. (2004). *Technical report: Cochlear implants* (ASHA Supplement 24, pp. 1–35). Rockville, MD: Author.

Best, C., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development, 21,* 295.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of the perceptual re-organization for speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345–360.

Bountress, N. G., Sever, J. C., & Williams, J. T. (1989). Relationship between two nontraditional procedures for assessing speech-sound discrimination. *Perceptual and Motor Skills, 69,* 499–503.

Chatfield, C. (2003). *The analysis of time series: An introduction* (6th ed.). Boca Raton, FL: Chapman & Hall/CRC.

Civan, A., Teller, D. Y., & Palmer, J. (2005). Relations among spontaneous preferences, familiarized preferences, and novelty effects: Measurements with forced-choice techniques. *Infancy, 7,* 111–142.

Cohen, L. B. (1969). Observing responses, visual preferences, and habituation to visual stimuli in infants. *Journal of Experimental Child Psychology, 7,* 419–433.

Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. *Infant Behavior and Development, 21,* 713–731.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (version 1.0) [Computer software]. Austin: University of Texas.

Cohen, L. B., DeLoache, J. S., & Rissman, M. W. (1975). The effect of stimulus complexity on infant visual attention and habituation. *Child Development, 46,* 611–617.

Cohen, L. B., & Oakes, L. M. (1993). How infants perceive a simple causal event. *Developmental Psychology, 29,* 421–433.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2,* 133–142.

Doehring, D. G. (1969). Temporal sequence effects in auditory oddity discrimination. *Perception and Psychophysics, 6,* 65–68.

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

Eisenberg, L. S., Martinez, A. S., & Boothroyd, A. (2004). Perception of phonetic contrasts in infants: Development of the VRASPAC. In R. T. Miyamoto (Ed.), *International congress series: Vol. 1273. Cochlear implants* (pp. 364–367). Amsterdam: Elsevier.

Horowitz, F. D. (1975). Infant attention and discrimination: Methodological and substantive issues. *Monographs of the Society for Research in Child Development, 39*(Serial No. 158).

Houston, D. M. (2005). Speech perception in infants. In D. B. Pisoni & R. R. Remez (Eds.), *Handbook of speech perception* (pp. 417–448). Cambridge, MA: Blackwell.

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

Houston, D. M., Ying, E., Pisoni, D. B., & Kirk, K. I. (2003). Development of pre word-learning skills in infants with cochlear implants [Monograph]. *The Volta Review, 103,* 303–326.

Hunter, M. A., & Ames, E. W. (1989). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 5, pp. 69–93). Norwood, NJ: Ablex.

Jusczyk, P. W. (1997). *The discovery of spoken language.* Cambridge, MA: MIT Press.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64,* 675–687.

Kagan, J., & Moss, M. (1965). Studies of attention in the human infant. *Merrill-Palmer Quarterly, 11,* 95–127.

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

Kirk, K. I., Miyamoto, R. T., Ying, E. A., Perdew, A. E., & Zuganelis, H. (2002). Cochlear implantation in young children: Effects of age at implantation and communication mode [Monograph]. *The Volta Review, 102,* 127–144.

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Perception* (Vol. 2, pp. 41–66). New York: Academic.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5,* 831–843.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period." *Language Learning and Development, 1,* 237-264.

Lewkowicz, D. J. (2000). Infants' perception of the audible, visible, and bimodal attributes of multimodal syllables. *Child Development, 71,* 1241–1257.

McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition, 95,* B15–B26.

Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development, 6,* 313–330.

Moore, J. M., Wilson, W. R., & Thompson, G. (1977). Visual reinforcement of head-turn responses in infants under 12 months of age. *Journal of Speech and Hearing Disorders, 42,* 328–334.

Newman, R. S., Bernstein Ratner, N., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology, 42,* 643–655.

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

Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: Different water in the same river. *Audiology and Neuro-Otology, 5,* 111–139.

Pisoni, D. B. (2000). Cognitive factors and cochlear implants: Some thoughts on perception, learning, and memory in speech perception. *Ear and Hearing, 21,* 70–78.
Pisoni, D. B., Cleary, M., Geers, A. E., & Tobey, E. A. (2000). Individual differences in effectiveness of cochlear implants in children who are prelingually deaf: New process measures of performance. *The Volta Review, 101,* 111–164.
Polka, L., & Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 421–435.
Primus, M. A. (1992). The role of localization in visual reinforcement audiometry. *Journal of Speech and Hearing Research, 35,* 1137–1141.
Purdy, S. C., Katsch, R., Dillon, H., Storey, L., Sharma, M., & Agung, K. (2004). Aided cortical auditory evoked potentials for hearing instrument evaluation in infants. In R. Seewald & J. Gravel (Eds.), *A sound foundation through early amplification: Proceedings of 3rd Pediatric Conference* (pp. 115–127). Chicago: Phonak.
Reynell, J. K., & Huntley, M. (1985). *Reynell Developmental Language Scales* (2nd ed.). Windsor, UK: NFER-Nelson.
Richard, J. F., Normandeau, J., Brun, V. R., & Maillet, M. (2004). Attracting and maintaining infant attention during habituation: Further evidence of the importance of stimulus complexity. *Infant and Child Development, 13,* 277–286.
Sarant, J. Z., Blamey, P. J., Dowell, R. C., Clark, G. M., & Gibson, W. P. R. (2001). Variation in speech perception scores among children with cochlear implants. *Ear & Hearing, 22,* 18–28.
Slater, A., Rose, D., & Morison, V. (1984). New-born infants' perception of similarities and differences between two- and three-dimensional stimuli. *British Journal of Developmental Psychology, 2,* 287–294.
Strange, W., Polka, L., & Dittmann, S. (1986). Training intraphonemic discrimination of /r/-/l/. *Bulletin of the Psychonomic Society, 24,* 419–422.
Thompson, R. F., & Spencer, W. A. (1966). Habituation: A model phenomenon for the study of neuronal substrates of behavior. *Psychological Review, 73,* 16–43.
Tobey, E., Geers, A. E., & Brenner, C. (1994). Speech production results: Speech feature acquisition. *The Volta Review, 96,* 109–130.
Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development, 75,* 1067–1084.
Werker, J. F., Shi, R., Desjardins, R., Pegg, J. E., Polka, L., & Patterson, M. (1998). Three methods for testing infant speech perception. In A. Slater (Ed.), *Perceptual development: Visual, auditory, and speech perception in infancy* (pp. 389–420). East Sussex, UK: Psychology Press.
Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49–63.