Four-Month-Olds’ Discrimination of Voice Changes in Multimodal Displays as a Function of Discrimination Protocol

Jason S. McCartney
Department of Psychology
Salisbury University

Robin Panneton
Department of Psychology
Virginia Tech

Past studies have found equivocal support for the ability of young infants to discriminate infant-directed (ID) speech information in the presence of auditory-only versus auditory + visual displays (faces + voices). Generally, younger infants appear to have more difficulty discriminating a change in the vocal properties of ID speech when they are accompanied by faces. Forty 4-month-old infants were tested using either an infant-controlled habituation procedure (Experiment 1) or a fixed-trial habituation procedure (Experiment 2). The prediction was that the infant-controlled habituation procedure would be a more sensitive measure of infant attention to complex displays. The results indicated that 4-month-old infants discriminated voice changes in dy-namic face + voice displays depending on the order in which they were viewed during the infant-controlled habituation procedure. In contrast, no evidence of discrimination was found in the fixed-trial procedure. The findings suggest that the selection of experimental methodology plays a significant role in the empirical observations of infant perceptual abilities.

The environments of young infants are filled with events that are comprised of various combinations of sensory inputs. On a typical day an infant hears voices,
manipulates objects, and crawls on surfaces, and in each of these situations the infant perceives the world with multiple sensory systems at the same time. Several developmental researchers have offered recent arguments for the importance of methods to reflect the rich, multisensory nature of the experiences that form the basis for many of our research questions about infant development (Dent-Read & Zukow-Goldring, 1997; Lewkowicz, 2001; Lickliter & Bahrick, 2001; Schmuckler, 2001; Walker-Andrews, 1997; Walker-Andrews & Bahrick, 2001). For example, when infants hear voices they also typically see lips opening and closing, heads nodding or turning, and eyes blinking. Thus, ecologically sensitive studies of infant speech perception would include corresponding visual as well as auditory information. Of primary interest here is that many of the studies that comprise the fields of infant visual and auditory perception have obtained results using methods that are designed to concentrate on one sensory modality at a time. That is, most studies tapping into infants’ visual preferences and discriminations have not included concurrent auditory input. Likewise, most studies tapping into infants’ auditory preferences and discriminations have not included concurrent visual information.

There is reason to believe that we cannot assume that patterns of perceptual responsiveness to different events remain constant independent of the arrangements under which infants are tested. One example comes from the literature on infants’ preferences for infant-directed (ID) speech. ID speech is characteristic of adults conversing with infants or toddlers and differs from adult-directed (AD) speech in that it is higher in pitch, has a wider pitch range, and has smooth and modulated pitch contours (Fernald & Simon, 1984; Papousek, Papousek, & Bornstein, 1985). In multiple studies in which infants are given the choice of hearing ID or AD speech, they often behave in ways that facilitate their access to ID speech (Cooper, Abraham, Berman, & Staska, 1997; Cooper & Aslin, 1990; Fernald, 1985; Papousek & Papousek, 1981; Pegg, Werker, & McLeod, 1992; Werker & McLeod, 1989). For the most part, these studies have examined infants’ attention to different speaking styles under impoverished ecological conditions (i.e., auditory-only manipulations). The ID or AD speech is often paired with static, nonfacial visual displays.

In an interesting series of experiments, Lewkowicz (1996, 1998) examined developmental differences in infants’ discrimination of both AD and ID vocal and facial information, but used testing conditions that were designed to improve ecological validity. In one set of studies (Lewkowicz, 1996), 4-, 6-, and 8-month-old infants were exposed to a film of an actor reciting a prepared script as if reading it to another adult (AD speech). Then, infants were given three test trials to determine to which aspect of the face and voice they had attended. The auditory-only (A) test trial consisted of the familiar face paired with a novel voice uttering a novel script. The visual-only (V) test trial consisted of a novel face paired with the famil-
iar voice uttering the familiar script. Finally, the combined test trial (AV) involved the presentation of a novel face and voice reciting a novel script (i.e., new actor). According to Lewkowicz, the A and V test trials examined infants’ abilities to discriminate modifications within a single sensory modality, and the combined (AV) test trials examined the infants’ abilities to discriminate changes in both modalities concurrently. The results of Lewkowicz’s (1996) first two experiments supported previous results with dynamic bouncing discs (Lewkowicz, 1992) in that the infants exhibited visual dominance across the separate test trials (i.e., they showed evidence of discrimination for changes in the V and AV test trials). More important, the infants (regardless of age) failed to discriminate the changes on the A trials. These findings are surprising given that (a) the lexical-syntactic content of the script changed, (b) speaker identity changed, and (c) the vocal and visual aspects of the face became desynchronized. Consequently, Lewkowicz (1996) suggested that his selection of speech type (AD speech) could have been insensitive to infants’ auditory abilities.

Thus, in his third experiment, Lewkowicz (1996) tested infants’ bimodal perception of ID and AD speech using two groups. To accentuate the difference between the ID and AD multimodal speech Lewkowicz filmed a female actor reciting script in an ID manner and filmed a male actor reciting script in an AD manner. Therefore, the infants were tested with the “greatest possible difference in type of speech” (Lewkowicz, 1996, p. 358). One group was exposed to multimodal ID displays and then tested with AD displays, whereas the second group was exposed to the reverse test order. The findings for the ID–AD group supported previous results with ID speech (Pegg et al., 1992). That is, all the ID-first infants spent a greater amount of time looking during the habituation phase when compared to the AD-first group. However, both the 4- and 6-month-old infants failed to show a significant decrease in attention to the multimodal ID displays over trials (i.e., they did not completely habituate to the display). Consequently, none of the 4- or 6-month-olds showed significant recovery of attention to the A, V, or AV test trials. In contrast, the 8-month-old infants showed significant recovery to all trial types (A, V, and AV). Lewkowicz suggested that the greater attractiveness of the ID speech might have made it more difficult for the younger infants to respond to any of the manipulations in the test trials.

The results of the AD-first group stand in sharp contrast to those of the ID-first group. All three age groups successfully habituated to the AD display and showed significant recovery of attention to the V and AV test trials. However, only the 6- and 8-month-olds exhibited significant recovery to the A test trial. Lewkowicz (1996) interpreted these findings as providing additional support for the dominance of the visual modality in infants’ bimodal perception of dynamic displays, particularly at younger ages. The 6- and 8-month-olds in the AD-first group were the only infants across all three studies who discriminated
the auditory change. This finding challenges the view that infants are differentially attentive to the vocal properties of their interactions with adults, in the absence of concurrent visual information, because most studies on infants’ preferences for ID speech have not included dynamic face + voice displays. In other words, to generalize that infants attend differentially to ID speech based on past studies is misleading because they appear to do so only when the visual information that is present is arbitrary (e.g., a bull’s-eye) and not ecologically linked to the speech itself (i.e., a dynamic face).

More recently, Lewkowicz (1998) found evidence that 3-month-olds detect vocal changes in a bimodal display when they were first habituated to a male, AD face + voice display, and then tested with a male face + female singing display. Singing is characteristically similar to ID speech because it involves exaggerated rhythms, pitch contour patterns, and stress patterns (Trehub, Unyk, & Kamenetsky, 1997). Lewkowicz suggested that the combination of gender-related cues and exaggerated prosody changes may have allowed even younger infants to make a successful auditory discrimination. The ability of young infants to discriminate a change from AD to ID singing but not AD to ID speaking is surprising, however, given that ID speech has also been called “musical speech” (Trainor, Clark, Huntley, & Adams, 1997) and that the “melody is the message” in ID speech to young infants (Fernald, 1989, 1992), making ID speech more akin to music early in development. On the other hand, it is possible that infants discriminated a change from AD to ID singing because of the ability of lullabies to convey more emotional intent to infants (Trehub, Unyk, & Trainor, 1993). Recent findings indicate that infants prefer speech that is rated as high in vocal affect, independent of whether the speech is ID or AD (Kitamura & Burnham, 1998; Singh, Morgan, & Best, 2002). Thus, it is possible that ID singing is perceptually more salient than ID speaking, making it easier for infants to notice changes in the voice even in the presence of a dynamic face. This implies that in natural interactions with caretakers that simply involve ID talking (not singing), younger infants are not as likely to attend to their voices.

However, another possible explanation for younger infants’ apparent insensitivity to changes in a single modality (particularly speech) could be due to the orchestration of certain habituation procedures. More specifically, Lewkowicz (1988a, 1988b, 1992, 1996, 1998) used a fixed-trial habituation design in which infants were given 12 trials of set duration. Infants in fixed-trial procedures have no control over the duration of stimulus presentation, and this lack of control highlights a potential flaw in producing habituation (Haaf, Smith, & Smitley, 1983; Horowitz, 1974; Millar & Weir, 1995). Infants in fixed-trial procedures often have their looks shortened by the standardized offset of the experimental display (Millar & Weir, 1995). That is, infants who are focused on the display at the end of the trial get cut off when the display terminates, resulting in the truncation of looking time (i.e., the
infant may have been still “attending” when the event was terminated). In addition, fixed-trial habituation procedures restrict the number of trials received by individual infants, essentially assuming that all infants habituate within the same number of exposures.

Horowitz (1974) discussed the drawbacks associated with the fixed-trial habituation procedure and suggested a possible alternative design, infant-controlled habituation. In her original work on infant visual attention, Horowitz noticed that infants exhibited differential patterns of looking (see also Colombo, 1993). For example, some infants would gaze at the event at the end of the trial only to be cut off and others would take many short glances before deciding to fixate for an extended period of time. Therefore, Horowitz and her colleagues decided to leave the display on for as long as each infant looked at it (Horowitz, Paden, Bhana, & Self, 1972). They found two interesting findings: Individual infants exhibited different durations of looks, and there was a greatly reduced percentage of subject loss.

Infant control habituation procedures allow individual infants to manage the duration of event exposure and permit researchers to bring every infant to the same relative level of response decrement (Horowitz, 1974). The classic infant control design involves using individual infants’ first several looks as the baseline level of initial attention against which habituation (response decrement) is measured for that infant. Subsequent looks are then compared to this average length of fixation. Typically, once the infant drops below 50% of the original average looking time, the habituation criterion is achieved.

As such, Lewkowicz’s (1996, 1998) design may have underestimated infants’ intersensory capabilities due to the missing contingent relation between infants’ active looking and the duration of event exposure. That is, some of the infants may not have had adequate time to process aspects of the face + voice displays. The goal of this study was to extend Lewkowicz’s (1996) findings by comparing fixed-duration and infant-controlled habituation procedures within the context of infants’ discrimination of ID face + voice displays.

The set of experiments reported here investigated 4-month-old infants’ abilities to discriminate specific auditory-only manipulations (e.g., gender of speaker’s voice and speech style: female ID vs. male AD) using an infant-controlled habituation procedure (Experiment 1) and a fixed-trial procedure (Experiment 2). We included both a change in gender and a change in speaking style to more directly compare our results to those of Lewkowicz (1996) in which 4-month-olds did not show discrimination between a male AD face/male AD voice and a male AD face/female ID voice display. The infant-controlled habituation procedure was expected to be more sensitive to infants’ individual variability in looking patterns (Horowitz, 1974) and reveal infants’ attention to ID speech in a multimodal context.
EXPERIMENT 1: VOCAL DISCRIMINATION IN AN INFANT-CONTROLLED PROCEDURE

Method

Participants. Twenty 4-month-old infants participated in Experiment 1 ($M = 129.3$ days, $SD = 6.1$; 14 girls and 6 boys). Of the 14 girls, 7 were assigned to one of the two testing orders; of the 6 boys, 3 were assigned to one of the two testing orders. The data of 8 additional infants were eliminated from the study (3 for fussiness, 4 for equipment failures, and 1 fell asleep). The parents of the infants were recruited for participation through local birth announcements. Parents were sent a letter describing the study, and subsequently contacted by phone to see if they were interested in participating in the study. Demographic information was obtained from each parent via a questionnaire given on the day of testing. The demographics of the final sample were as follows: 100% were White, 100% were from married homes, 80% were delivered vaginally, 80% were breastfed, and 45% were first-born. The average age of the infants’ mothers was 31 ($SD = 5.70$). The socioeconomic status of the families was middle to upper middle class.

Audiovisual displays. There were four audiovisual films used in this experiment, two for the habituation phase and two for the testing phase. All the films were made using a Sony Hi8 video camera (model CCD-TRV72) and an Adobe Premiere 5.0 video editing computer program on an IBM PC 300 PL. In addition, all the audiovisual editing was performed with a Panasonic editing controller (model AG-A770P), a Panasonic digital audiovisual mixer (model WJ-MX12), and a professional/industrial Panasonic videocassette recorder (model AG-7750). The two films used in the habituation phase consisted of a male or a female actor reciting a selected portion of the children’s story Santa Mouse1 with the following differences: The man recited the story in an AD speaking style, whereas the woman recited the story in an ID speaking style. All infants received synchronous and matching face + voice (i.e., male face + male voice) displays during habituation.

The four films that were used during the test trials (i.e., after the habituation criterion was met) also contained audiovisual information, but the voice tracks were switched (face tracks remained the same). For example, the infants habituated to the female ID face + voice film ($n = 10$) received the familiar female ID face paired with the novel male AD voice. Similarly, the infants habituated to the male AD

---

1Words from Santa Mouse that were recited for the films: “Once there was a little mouse who didn’t have a name. He lived in a great big house, this mouse. The only mouse in the whole, wide house. On Christmas Eve, he brushed his teeth, and as he washed his paws, he thought, ‘My goodness! No one gives a gift to Santa Claus.’” Average length from start to finish was 15.09 sec (male AD) and 15.84 sec (female ID).
face + voice film \((n = 10)\) received the familiar male AD face paired with the novel female ID voice. The test films were constructed using an Adobe Premiere 5.0 video editing computer program, which allowed the researcher to manipulate the exact timing of the auditory and visual displays. To create synchronous films, both speakers recorded a second soundtrack of their ID (female) and AD (male) recitations while viewing their own as well as the other speaker’s film and synchronizing their voice to the corresponding lip movements in each. That is, they matched (in real time) the periods of voicing and nonvoicing with the mouth movements of themselves and the other speaker. The films were then created by lining up one of the voice tracks (e.g., female ID) with one of the visual tracks (e.g., male AD). Lining up these two tracks was accomplished both by visual inspection of the simultaneous tracks in the Adobe Premiere program, and also by inserting markers at the specific periods of voicing and nonvoicing on the video tracks to ensure that a voiced segment of the video was accompanied by a voiced segment of the audio track. In other words, every utterance was matched up with the speaker’s facial movements corresponding to each segment of the text (i.e., whenever the speaker’s mouth opened and closed).

We also created asynchronous versions of each film, constructed in much the same way as just discussed, except that the second speaker’s film started at a different point in the script compared to the first speaker. That is, one speaker’s lip and head movements failed to correspond to the same movements in the other speaker. This was accomplished by offsetting the video and audio tracks by approximately 5 sec (male video/female audio) and 4.5 sec (female video/male audio). Half of the infants \((n = 10)\) received a synchronous film during the test phase, whereas the other half \((n = 10)\) received an asynchronous film during the test phase (again, all infants received synchronized face + voice displays during habituation). However, because synchrony status did not emerge as a significant variable with respect to infants’ attention or discrimination of vocal tracks, it was dropped from all further analyses of the data.\(^2\)

**Apparatus.** Infants were tested while sitting on their parent’s lap in the experimental room. The parent was seated in a chair facing a black wooden three-sided enclosure, 80 cm (length) × 80 cm (width) × 60 cm (height). To the parent’s right and rear was a white wall, and to his or her left was a black covering (piece of cardboard). This covering served to restrict the infant’s peripheral view. Separating the parent and the infant from the front panel was a 40 cm × 80 cm wooden shelf painted black and covered with a white foam pad. This shelf pro-

\(^2\)The amount of asynchrony between the faces and voices in both the male and female films was above the level that has been determined to be sufficient for infant detection (Lewkowicz, 2000). Thus, the lack of effect of synchrony on infant discrimination was due to some factor other than a lack of detection.
vided the infant with support and a safe place to touch during the experimental session. A 27.5 cm × 21 cm television monitor (RCA model E13209 BC) was positioned behind an opening in the front panel of the enclosure offset 7.6 cm to the right of midline, approximately 35 cm away from the infant. A small speaker (Jamo compact 60) was also located directly below the television screen. A video camcorder (Panasonic model AG-180) was also behind the front panel with its lens positioned behind a hole (3.5 cm radius) in the front panel. The camcorder recorded the infants’ responses and provided a view of each infant to an observer.

The observer utilized a Power Mac 7100/80 computer to control the presentation of the audiovisual films, control time between stimulus presentations, and store the length of infant visual fixations. The observer controlled the presentations of the visual display, controlled the speech films, and recorded the individual looking times via a handheld microswitch attached to the computer. Attached to this computer was also a custom-built control board and two videocassette player/recorders (Sanyo model VHR-5214; Panasonic model AG-1960). All the videotapes of the male and female actors were played on one of these two VCRs and were presented via the television monitor to the infant. The audio output was sent through an amplifier (Harmon-Kardon model PM635) and presented through the panel speaker at 70 to 75 dB SPL (A-scale; measured at the infant’s head) as determined by a sound-level monitor (Radio Shack, Cat. No. 33–2050).

The observer was unable to hear the stimuli due to continuous, loud vocal music (delivered over headphones) and viewed the infant on an 11.5 cm × 9 cm television monitor (Sony Solid State model 21 CFR) during testing (this monitor was connected to the camcorder that broadcast the infant’s face). A small red light in front of the observer (attached to the top of the monitor) indicated when the infant’s visual display was on.

Procedure and design. The procedure selected for this experiment was a modified version of the infant-controlled habituation procedure (Horowitz, 1974). Once the infant was judged to be awake, the habituation phase of experimental testing began. The first trial started when the observer judged that the infant was looking at the television monitor and signaled the computer by pressing a button. Each trial continued until the infant was judged to be looking away from the screen for at least 1.5 sec (there was no maximum trial length). When the infant looked away, the video screen went black, and the observer waited until the infant was generally centered (i.e., looking toward the midline of the display area) before initiating the next trial. Once the video display was available, the observer would press the button whenever the infant was judged to be looking at the screen. Half of the infants were habituated to the male display, whereas the other half were habituated to the female display. Each infant’s mean looking time to the film across the first three trials (Trials 1–3) was computed and stored by the computer for later reference. This mean looking time was then compared to the average of every three
consecutive trials (starting with Trials 4–6). When this running average dropped to 50% or less of the mean of the first three trials, the habituation criterion was met. Thus, the minimum number of trials that all infants had to complete was six; however, there was no limit to the number of trials that could be experienced by any one infant. All sessions continued until the habituation criterion was met.

Following habituation, all of the infants (i.e., male first and female first) received two no-change posthabituation trials (hereafter called lag trials) to allow for the assessment of spontaneous regression effects (Bertenthal, Haith, & Campos, 1983). Because infants often exhibit attention patterns in which long looks follow short looks or vice versa, changes in infant looking time may be unrelated to the change in the stimulus. If this were the case, infants could look longer on the lag trials even though the auditory stimulus did not yet change. Following the two lag trials each infant received four test trials to assess visual recovery using one of the test tapes. Infants habituated to the female ID movie received the familiar woman’s face paired with the man’s AD voice. Likewise, the infants habituated to the male AD movie received the familiar man’s face paired with the female’s ID voice. The dependent variable was the total amount of looking time during each test trial.

Interrater reliability between the original coder and one offline coder was obtained for 20% of the participants by having the offline coder watch videotapes of these sessions, and record the onset and offset of each individual trial with no soundtrack available (the observer was also blind to the ordering of the displays). It should be noted that given the nature of the testing protocol, it was apparent when the video screen was on because the infant’s face would increase in illumination (emanating from the screen itself). However, the actual trial did not commence until the infant was judged to be looking at the screen, so there should not have been any clues to this onset time for the offline coder. Likewise, the end of a trial was also apparent because the illumination on the infant’s face would decrease. In this case, a cue may have been provided to the offline observer because this coder would know that the original observer had terminated the trial, potentially inflating reliability estimates. To deal with this potential bias, we calculated reliability in two ways: a bivariate correlation between all trials \((n = 56)\) and a second correlation only between trials in which the duration was longer according to the original coder (i.e., the offline coder determined that the infant had looked away prior to the time that the original coder did so; \(n = 36\)). The respective correlations were \(r = .99\) for each comparison \((p < .01, \text{two-tailed})\). The mean difference in looking times per trial across the two observers was 0.35 sec \((SD = 1.32)\).

Results and Discussion

As a preliminary analysis, we examined whether infants looked longer during the initial three trials (the average of which served as the habituation criterion) as a function of whether they were presented with a male AD face + voice or a female
ID face + voice display. Despite longer mean looking times during baseline to the 
AD male face + voice, this difference was not statistically significant ($M_{\text{male}} = 
41.24 \text{ sec}, SE = 12.01; M_{\text{female}} = 26.42 \text{ sec}, SE = 3.99), t(18) = 1.71, p > .05. Given 
that the mean looking times during baseline were more variable in the male face + 
voice group, we looked at individual infants’ data and determined that 1 infant in 
this group had a mean baseline of 116.8 sec, 2 standard deviations above the group 
mean. When this infant’s looking time is removed, the mean baseline for this group 
is 32.84 ($SE = 9.59), more in line with that for the female face + voice group. We 
also compared the average number of trials that it took infants in each of the two 
orders to reach our criterion for habituation, but these were also not significantly 
different ($M_{\text{male}} = 14.6 \text{ trials}, M_{\text{female}} = 10.6 \text{ trials}), t(18) = 0.24, p > .05.

To test for vocal discrimination, infants’ mean looking times were compared in a 
mixed analysis of variance (ANOVA), with order (female ID first, male AD first) as 
the between-subjects factor, and trials (mean lag, mean post) as the within-subjects 
factor. The average of the two lag trials was used (rather than the average of the three 
trials that met the habituation criterion) to control for spontaneous recovery in visual 
attention. The mean post scores were the average of the first two trials after these lag 
trials, during which a change in voice was available. The results of this analysis 
showed nonsignificant main effects for both order, $F(1, 18) = 2.93, p > .05$, and trials, 
$F(1, 18) = 3.11, p > .05$, but a statistically significant Order × Trials interaction, $F(1, 
18) = 6.09, p < .03$. Analysis of this interaction showed that infants had significantly 
longer average looks on the post trials ($M = 27.93 \text{ sec}, SE = 7.82$) compared to their 
lag trials ($M = 14.71, SE = 4.15$) when they were presented with the AD male face + 
voice first, and then tested with the AD male face + ID female voice, $t(9) = 2.36, p < 
.05$ (Cohen’s effect estimate $d = .70$). In contrast, they did not look significantly lon-
ger on post trials when presented with the ID female face + male AD voice ($M = 10.37 
\text{ sec}, SE = 1.65$) after habituating to the ID female face + voice film ($M = 12.58 \text{ sec}, SE 
= 2.13), t(9) = 0.80, p > .05$ (Cohen’s effect estimate $d = .34$; see Figure 1).

The results of Experiment 1 show that 4-month-old infants discriminated 
changes in speakers’ voices without corresponding changes in the speakers’ faces, 
provided that the change was from a male AD to female ID voice. Infants did not 
show recovery of attention to the reverse order of these films. It is not uncommon 
for infants to show such asymmetry in recovery of attention after habituation, often 
if one of the events is thought to be more perceptually salient than the other. That 
is, events that are generally preferred by infants (i.e., have higher interest value; 
Lewkowicz, 1996) tend to generate more attentional recovery after habituation 
than events that are low in interest value. For example, after habituating to a picture 
of a sad emotion expression, 3-month-olds recovered attention to a picture of a sur-
prise emotion, but showed no recovery of attention in the reverse order 
(Walker-Andrews & Lennon, 1991; Young-Browne, Rosenfeld, & Horowitz, 
1977). Likewise, 4-month-olds showed discrimination of a surprise face after ha-
ituating to a happy face, but not in the reverse order (Caron, Caron, & Myers,
and 7-month-olds showed asymmetrical discrimination of a fearful versus a happy face (Nelson, Morse, & Leavitt, 1979). In terms of auditory discrimination, newborn infants showed increases in sucking activity to a female film of their native language after habituating to a female film of a nonnative language, but did not increase sucking (i.e., showed no discrimination) to the reverse presentation order (Mehler et al., 1988), presumably because even newborns prefer films of their native over nonnative languages (Moon, Cooper, & Fifer, 1993).

In terms of explaining the results reported here, it has been found repeatedly that infants prefer ID to AD speech (Cooper et al., 1997; Cooper & Aslin, 1990; Fernald, 1985; Pegg et al., 1992; Werker & McLeod, 1989). Although it is not clear if infants prefer female to male voices, several studies have shown differential responding to the maternal voice (Burnham, 1993; DeCasper & Fifer, 1980), especially maternal ID over AD speech (Cooper et al., 1997). In contrast, young infants do not prefer the paternal voice or find male voices reinforcing (DeCasper & Prescott, 1984; Ward & Cooper, 1999). However, at least one study found preferences for male ID over male AD speech (Werker & McLeod, 1989). For the 4-month-olds in this study, the change from AD male face + voice to ID female face + voice appears to have been more perceptually salient than the other way around. This is consistent with previous studies showing increased attention to ID speech, especially in younger infants.

Our finding that 4-month-olds discriminated a change in voice without an accompanying change in facial information is inconsistent with Lewkowicz’s (1996) results in that the infants in his study did not discriminate changes in the auditory modality alone without corresponding changes in the visual modality (unimodal
auditory discrimination was only seen at 6 months and older). Interestingly, he also found differential attention depending on order of presentation, with the 6- and 8-month-old infants receiving AD → ID order showing significant recovery, but not those in the opposite order. Unlike in this study, the 4-month-old infants in Lewkowicz’s (1996) study showed no significant response recovery in either order, even when the voice changed from that of a man to a woman. In a later study, Lewkowicz (1998) did find significant discrimination of just auditory information in 3-month-old infants when a male AD voice was changed to a singing female voice. It is possible that this discrepancy between our results and those of Lewkowicz (1996) was due to differential encoding opportunities between habituation protocols rather than differences in the properties of the face + voice displays per se. We assessed this possibility by conducting a second experiment that utilized a habituation methodology akin to Lewkowicz’s (1996, 1998) fixed-trial procedure.

**EXPERIMENT 2: VOCAL DISCRIMINATION IN A FIXED-TRIAL PROCEDURE**

**Method**

**Participants.** Twenty 4-month-old infants participated in Experiment 2 ($M = 130.9$ days, $SD = 4.7$; 12 boys and 8 girls). Of the 12 boys, 4 were assigned to the male AD face + voice first condition, and 8 to the female ID face + voice first condition. Of the 8 girls, 5 were assigned to the male AD face + voice first condition, and 3 to the female ID face + voice first condition. The data of 5 additional infants were rejected from the study (2 for excessive fussiness, 2 for equipment failure, and 1 for a sibling intrusion). The demographics of the final sample are as follows: 95% were White, 5% were African American, 100% were from married homes, 95% were delivered vaginally, 85% were breastfed, and 35% were firstborn. The average age of the infants’ mothers was 30 ($SD = 4.74$). The socioeconomic status of the families was middle to upper middle class.

**Audiovisual displays.** The audiovisual films used in Experiment 1 were again utilized in this experiment.

**Apparatus.** The apparatus was the same as that used in Experiment 1.

**Procedure.** Because we were primarily interested in infants’ abilities to discriminate a change in vocal information under conditions of viewing a concurrent speaking face, the infants were tested using a modified (i.e., truncated) version of Lewkowicz’s (1996, 1998) fixed-trial procedure. Infants in this study received one critical test trial (auditory alone), whereas infants in Lewkowicz’s multimodal research received three test trials (A, V, and AV). In addition, Lewkowicz included a
familiar trial (randomly interspersed among test trials) during which the infants were re-presented with the habituation display. He then compared responding on this trial with change trials for evidence of discrimination. Because we did not include multiple test (change) trials, we used performance on Trial 12 (the last familiarization trial) as our primary comparison for the test trial. Therefore, every infant in this experiment received twelve 20-sec trials in the habituation phase (i.e., male AD face + male AD voice) and one 20-sec A trial (e.g., male AD face + female ID voice) in the test phase (13 trials in total). All 13 fixed trials were separated from each other by a blank and silent screen for 3 sec.

Once the infant was judged to be awake and alert, the habituation phase of the experiment began when the observer pressed the key signaling the computer to start the procedure (Trials 1–13). On every 20-sec trial the film was presented continuously, regardless of each individual infant’s looking pattern (i.e., the movie with soundtrack remained on during the entire 20-sec interval). Then, at the end of a trial the television monitor and auditory channel (i.e., speaker) were turned off by the computer. Three seconds later the next trial began, until all 13 trials were complete. For every display trial, the observer recorded infants’ looking times by pressing the signal button. That is, during every 20-sec trial infants could exhibit several separate looks at the display. Consequently, the observer pressed the button once to signal the onset of a look and then again to mark the end of a look. The computer stored distinct looks by registering the trial number on which it occurred, the duration of the look, and the auditory channel being presented.

In terms of interrater reliability estimation, it was apparent when the video screen was on because the infant’s face would increase in illumination (emanating from the screen itself), as it did in Experiment 1. However, given that the infant could issue several looks within one trial in this procedure, the only time that a cue may have been provided to the offline observer regarding the onset and offset of individual looks was when the trial terminated (i.e., the illumination on the infant’s face decreased). To deal with this potential bias, we calculated reliability in two ways: a bivariate correlation between all looks (n = 66) and a second correlation only between looks prior to the termination of the trials (n = 44). The respective correlations were $r = .98$ and $r = .96$ (both $p < .01$, two-tailed). The mean difference in looking times per trial between the two observers was 0.36 sec ($SD = 1.71$).

Results and Discussion

To test for auditory discrimination, infants’ mean looking times were compared in a mixed ANOVA with order (male AD face + voice first, female ID face + voice first) as the between-subjects factor and trial (Trial 12 [prechange] vs. Trial 13 [postchange]) as the within-subjects factor. The results of this ANOVA did not show any significant main effects of order, $F(1, 18) = 0.75$; trial, $F(1, 18) = 0.83$; or the Order × Trial interaction, $F(1, 18) = 0.15$, all $ps > .05$. The mean looking times
from Trial 12 to Trial 13 for the AD male face + voice first group were 11.62 sec versus 12.16 sec (Cohen’s effect estimate $d = .09$), and those for the ID female face + voice group were 13.38 sec versus 14.7 sec (Cohen’s effect estimate $d = .23$), respectively. More important, we did not include an additional posttest trial during which the infants were given a completely novel display to rule out fatigue as the primary reason for their lack of recovery (see Lewkowicz, 1996, 1998). Despite this omission, it seems likely that fatigue was not pervasive given that the mean looking time on Trial 1 ($M = 14.22$ sec, $SE = 0.54$) was quite similar to the mean looking times on Trial 12 ($M = 12.5$ sec, $SE = 1.25$) and Trial 13 ($M = 13.45$ sec, $SE = 1.43$), suggesting that fatigue was not responsible for the lack of recovery to the vocal change.

Additional analyses were conducted to compare infant habituation patterns across experimental methodologies (i.e., infant controlled vs. fixed trial). One ANOVA compared average looking times on the first trial of each session, with order (male AD face + voice first, female ID face + voice first) and procedure (infant controlled, fixed trial) as between-subjects factors. This analysis revealed a significant main effect of procedure, $F(1, 36) = 6.59, p < .02$, with infants looking longer on Trial 1 of the infant-controlled procedure ($M = 42.76$ sec, $SE = 11.93$) than of the fixed-trial procedure ($M = 14.22$ sec, $SE = 0.54$). However, there was also a significant Procedure × Order interaction, $F(1, 36) = 3.92, p = .05$, with infants looking significantly longer on Trial 1 when presented with the male AD face + voice speech in the infant-controlled procedure ($M = 64.87$ sec, $SE = 22.04$) compared to the fixed-trial procedure ($M = 14.35$ sec, $SE = .71$).

In addition, to compare the procedures more directly, an ANOVA was conducted on the looking times on the last trial before the switch in displays occurred (Lag 2 for infants in Experiment 1; Trial 12 for infants in Experiment 2) compared to the first trial after the switch (Post 1 for infants in Experiment 1; Trial 13 for infants in Experiment 2), and as a function of order (male AD face + voice first, female ID face + voice first) and procedures (infant controlled, fixed trial). The results showed a statistically significant Procedure × Order × Trial interaction, $F(1, 36) = 3.82, p = .05$. This interaction complements the results of the two independent experiments in that infants looked significantly longer on the trial after the switch only in the male AD face + voice first condition, and only in the infant-controlled procedure (see Figure 2).

In contrast to the 4-month-olds in Experiment 1, there was no evidence that 4-month-old infants in Experiment 2 successfully discriminated the change in gender and speaking style of an adult’s speech. The secondary analyses involving the interaction between procedure and test trials found that only infants in the infant-controlled habituation procedure showed significant discrimination when presented with an auditory change. One possible explanation for this differential demonstration of discrimination is that the infants in Experiment 1 were allowed to attend to and encode the audiovisual display with no time constraint (i.e., individ-
ual trials terminated only when the infant’s attention terminated). In support of this interpretation, an analysis of the average length of habituation time between procedures showed that infants spent more time looking at the display in Experiment 1 \( (M = 368.52 \text{ sec}) \) than in Experiment 2 \( (M = 204.21 \text{ sec}) \), \( t(38) = 2.85, p < .05. \) Thus, the infants in the infant-controlled procedure had a higher likelihood of learning about the audiovisual display during habituation. Interestingly, this increase did not seem to be due to more trials in Experiment 1 because the mean trial number was 12.6 \( (SD = 7.51) \), quite close to the 12 fixed trials in Experiment 2. Rather, the encoding advantage was due to self-initiated and self-terminated attention during trials, and the longer overall looking times during the pretest phase.

**GENERAL DISCUSSION**

Results of two experiments using different methodologies found conflicting evidence for the ability of 4-month-old infants to discriminate vocal changes within the context of dynamic multimodal displays. Specifically, infants were tested for their abilities to discriminate a change in voice that involved both gender (i.e., male vs. female) and speaking style (i.e., AD vs. ID). Discrimination was only evident within the context of the infant-controlled habituation procedure (Experiment 1), and only in one of two presentation orders (going from a male AD face + voice display to a female ID face + voice display). No evidence for discrimination in either order was obtained when using a fixed-trial procedure (Experiment 2). A crucial
difference between the fixed-trial and the infant-controlled habituation procedure is individual control, manifested in infants’ looking patterns (Horowitz, 1974). Infants tested in the infant-controlled procedure are brought to the same relative level of response decrement prior to the presentation of the test trials. In contrast, infants in the fixed-trial procedure receive a preset number of exposure trials regardless of how long they attend to the displays.

The evidence for discrimination of vocal information from this investigation supports earlier work in our lab in which we found (using the infant-controlled habituation procedure) infants’ discrimination of filtered ID versus AD speech (Cooper & Aslin, 1994), male voices (Ward & Cooper, 1999), fundamental frequency contours (McCartney & Cooper, 1999), and linguistic content (Theaux, 1998). However, the results from Experiment 1 must be qualified with a consideration of presentation order in that the 4-month-olds only showed response recovery to the change in voice after habituating to a male AD face + voice and then receiving a male AD face + female ID voice; the reverse order did not produce significant response recovery. One interpretation of this asymmetry in recovery of attention relates to the well-established finding that young infants prefer ID to AD speech (although examined primarily under visually static, auditory-change-only test conditions; Cooper et al., 1997; Cooper & Aslin, 1990; Fernald, 1985; Fernald & Kuhl, 1987; Pegg et al., 1993; Werker & McLeod, 1989). Strong preferences for one of two events used during habituation or dishabituation procedures can lead to recovery of attention in one order (nonpreferred → preferred) but not the other (preferred → nonpreferred). Although not as much evidence exists regarding infants’ preferences for female over male voices, it has been found that young infants do not prefer male voices over silence (DeCasper & Prescott, 1984), prefer a film of an intrauterine heartbeat over a male voice (Panneton & DeCasper, 1984), and do not show preferences for their fathers’ faces, voices, or combined faces and voices, even though same-aged infants do prefer their mothers’ faces and voices (Ward, 1998; Ward & Cooper, 1999). Thus, the current order effect seems reasonable given that the male AD face + voice→male AD + female ID voice condition represented infants’ preferred vocal event.

The finding here that 4-month-olds discriminate a change in vocal information in the context of a dynamic speaking face stand in contrast to Lewkowicz’s (1996) results; he found that 4-month-olds did not show discrimination of a vocal change under similar testing conditions, but with a fixed-trial procedure (similar to our results in Experiment 2). However, Lewkowicz (1998) presented more current findings that young infants can discriminate changes in auditory-only information, but only under highly salient conditions (e.g., ID singing). In this more recent study, he found that 3-month-olds discriminated a change from a male AD face + voice to a male AD face + female singing voice (no synchrony) in the context of the fixed-trial procedure. Thus, it is certainly possible to find positive evidence of infants’ multimodal discrimination of various events with the fixed-trial procedure;
however, our current data suggest that a lack of evidence for discrimination with this procedure needs to be interpreted cautiously.

In addition, Lewkowicz’s (1996, 1998) interpretation that younger infants’ lack of discrimination of vocal changes within the context of dynamic visual and auditory displays reflects an insensitivity to auditory information is somewhat contradicted by work with older children. For example, Sloutsky and Napolitano (2003) recently argued for auditory dominance in 4-year-old children. Sloutsky and Napolitano presented compound static (not dynamic) visual and auditory stimuli to children and adults and found that children more readily encoded auditory information over visual information. That is, when auditory and visual stimuli were presented simultaneously, the children showed greater attention to the auditory component. The adult participants showed the exact opposite behavior, relying almost solely on the visual components. Sloutsky and Napolitano concluded that young children (unlike adults) “may be strongly biased to pay more attention to the auditory modality” (p. 831). This finding for increased processing of auditory information has been recently replicated using different auditory events, and also extended to 8-month-old infants (Robinson & Sloutsky, 2004). It is important to note, however, that these same outcomes may not be achieved if the visual stimuli are dynamic (i.e., moving) rather than static (see Lewkowicz, 1992).

In sum, the results of the experiments presented here emphasize that research protocols used to study infant perception are not all equal in their opportunities for information processing. For some perceptual contrasts (perhaps ones that are quite easy), this is inconsequential. For others, using a procedure that limits information processing time may underestimate infants’ abilities to perceive and encode information necessary to show evidence of discrimination or preference. The results of our experiments are a case in point; the infant-controlled procedure resulted in 4-month-olds exhibiting auditory discrimination, whereas the fixed-trial procedure resulted in 4-month-olds’ lack of discrimination of the same events. Infant-controlled habituation procedures allow infants to attain similar levels of response decrement prior to testing for specific discriminatory abilities. Each infant in Experiment 1 determined the number and duration of test trials presented and this directly determined the length of the habituation phase. In contrast, fixed-trial habituation procedures (Experiment 2) involve a predetermined number of trials of a specific duration. Infant behavior or attention plays no operational role during the habituation phase. Moreover, the infants are not required to exhibit a standardized level of habituation (i.e., 50%) to proceed to the testing phase. It is possible that infants in fixed-trial procedures are more likely to become familiarized to events, but not habituated. Familiarization and habituation can lead to different patterns of behavior, depending on the age of the infant, complexity of the stimuli, and the number of presentations (Cohen, 2001). The use of both fixed-trial and infant-controlled habituation procedures has produced interesting and important findings regarding developmental changes in infants’ perception of events. None-
theless, the failure to find evidence of discrimination places more demand on the internal validity of these procedures if such findings are to be integrated into general statements about perceptual limitations during infancy. In this regard, the infant-controlled procedure may be the more appropriate one when the events to be discriminated require increased processing resources from infant perceivers.

ACKNOWLEDGMENTS

This research was submitted by Jason S. McCartney in partial fulfillment of the PhD requirements at Virginia Tech. Portions of this research were presented at the International Conference on Infant Studies, Toronto, Ontario, Canada, April 2002. We thank the graduate and undergraduate students at the Virginia Tech Infant Perception Lab for their help during data collection, and we acknowledge the assistance of the parents and children who participated in this study.

REFERENCES


Lewkowicz, D. J. (2001). The concept of ecological validity: What are its limitations and is it bad to be invalid? Infancy, 2, 437–450.


