Developmental Psychology

Infant biases for detecting speech in complex scenes

--Manuscript Draft--

Manuscript Number:	DEV-2019-2627R1
Article Type:	Article
Full Title:	Infant biases for detecting speech in complex scenes
Abstract:	How do infants learn the sounds of their native language when there are many simultaneous sounds competing for their attention in the real world? Adults and children detect when speech changes in complex scenes better than changes of other sounds. We examined whether infants have similar biases to detect changes to human speech better than non-speech sounds including music, water, and animal calls in complex auditory scenes. We used a change deafness paradigm to examine whether 5-month-olds' change detection is biased toward certain sounds within high-level categories (e.g., biological, or generated by humans) or whether change detection depends on low-level salient physical features and thus is better for more acoustically distinct sounds, such as water. In Experiment 1, 5-month-olds showed some evidence of detecting when speech and music sounds change better than no change trials. In Experiment 2, when speech and music were compared separately with animal and water sounds, infants detected when speech and water changed, but not when music changed across scenes. Infants' change detection is both biased for certain sound categories, as they detected small changes to speech better than other comparable sounds, and biased by the size of the acoustic change, similar to young infants' attentional priorities in complex visual scenes. By 5 months, infants show some preferential processing of speech changes in real-world auditory environments which could help bootstrap the language learning process.
Keywords:	Change deafness; infant change detection; attentional bias; speech; category knowledge; perceptual salience

April 5, 2019

Dear Dr. Dubow,

Thank you for reviewing our manuscript and providing some feedback inquiring about how we determined out sample sizes. We have now edited the manuscript to reflect this process, which included looking at previous studies' sample sizes and estimating our sample size a priori using G*Power. These changes are reflected in footnotes to the participants section in the method on pages 9 and 17.

Thank you,

The Authors

Abstract

How do infants learn the sounds of their native language when there are many simultaneous sounds competing for their attention in the real world? Adults and children detect when speech changes in complex scenes better than changes of other sounds. We examined whether infants have similar biases to detect changes to human speech better than non-speech sounds including music, water, and animal calls in complex auditory scenes. We used a change deafness paradigm to examine whether 5-month-olds' change detection is biased toward certain sounds within high-level categories (e.g., biological, or generated by humans) or whether change detection depends on low-level salient physical features and thus is better for more acoustically distinct sounds, such as water. In Experiment 1, 5-month-olds showed some evidence of detecting when speech and music sounds change better than no change trials. In Experiment 2, when speech and music were compared separately with animal and water sounds, infants detected when speech and water changed, but not when music changed across scenes. Infants' change detection is both biased for certain sound categories, as they detected small changes to speech better than other comparable sounds, and biased by the size of the acoustic change, similar to young infants' attentional priorities in complex visual scenes. By 5 months, infants show some preferential processing of speech changes in real-world auditory environments which could help bootstrap the language learning process.

Keywords: Change deafness, infant change detection, attentional bias, speech, category knowledge, perceptual salience

Introduction

Infants are born into a rich auditory environment of sounds that overlap and compete with each other. Prioritizing particular sounds in a real-world environment is critical for processing and learning about those sounds (Vouloumanos & Waxman, 2014). A critical early step for building infants' language system is to selectively attend to speech sounds even when other sounds are presented simultaneously. Adults and children as young as 6 years old are better at encoding and detecting changes involving the human voice in complex auditory scenes (Vanden Bosch der Nederlanden, Snyder, & Hannon, 2016). However, no study to date has examined whether infants have a processing advantage for detecting speech when presented simultaneously with competing sounds. We examined whether processing biases for detecting speech in complex auditory scenes are evident during infancy, when they could provide a potential mechanism for facilitating language learning.

Infants' early and powerful speech perception abilities have been explored through decades of innovative laboratory studies, including investigations of speech preference (e.g., Vouloumanos & Werker, 2004; 2007), discrimination (Kuhl, 2004; Werker & Tees, 1984; Werker, Yeung, Yoshida, 2012), segmentation (Saffran, Aslin, & Newport, 1996; Jusczyk, Houston, & Newsome, 1999; Thiessen, Hill, & Saffran, 2005), and categorization (Waxman, Markow, 1995; Werker et al., 1998; Swingley, 2009; Lim, Lacerda, & Holt, 2015). Laboratory studies, however, are typically conducted with speech presented in quiet, even sound-attenuated, conditions. In our daily interactions, human speech is often heard overlapping with other sounds, like dogs, cats, airplanes, construction, wind, or rain, which also compete for our attention (Bregman, 1990). Very few studies have examined speech perception in the noisier ecological conditions more typical of human experience (cf., Newman & Jusczyk, 1996; Newman et al., 2006). Becoming an expert at speech perception requires infants to selectively attend to human communicative sounds in auditory scenes that have multiple sounds competing for their attention.

Attention allocation in complex scenes is a critical and well-studied aspect of visual scene processing (e.g., Mitroff, Simons, & Levin, 2004; Simons & Rensink, 2005). In the visual domain, infants are biased to attend toward face or face-like stimuli from birth (e.g., Farroni et al., 2005; Simion et al., 2001) and infants' bias for attending to faces in dynamic visual scenes increases with age during their first year (Frank, Vul, & Johnson, 2009). Visual scene processing studies also highlight some of the factors that influence how infants allocate their attention. For example, younger infants of 3 months attend more to low-level salient physical features compared to high-level and semantically meaningful combinations of features specific to faces (Frank, Vul, & Johnson, 2009), suggesting it may take time for infants to fully develop a bias for faces *qua* faces over salient changes within a visual scene. Perceptual biases for faces seem to correspond with better processing of faces even in complex visual scenes. Specifically, both children and adults are faster at detecting a change in human faces than in environmental objects in complex scenes (Fletcher-Watson et al., 2009; Ro, Russel, & Lavie, 2001) and this advantage for face processing is seen in infants as young as 6 months (Gliga et al., 2009). Specific, early emerging biases may contribute to developing expertise in the perception of evolutionarily and socially important aspects of the visual world like faces.

In the same way that infants have attentional biases for faces, infants have early emerging biases for evolutionarily and socially relevant sounds such as speech. Even newborns listen longer to speech sounds compared to carefully matched non-speech stimuli in silent, controlled environments (Vouloumanos & Werker, 2004; 2007). Infants' listening bias for speech becomes more specific over the first 3 months, with infants preferring to listen to speech over monkey calls, and other human-produced non-speech sounds, like throat clearing or laughing (Shultz & Vouloumanos, 2010). Just as they have a processing advantage for faces, older children and adults have processing biases for speech in complex auditory scenes: they detect changes speech sounds more accurately than non-speech sounds when as many as four competing sounds are presented at the same time (Vanden Bosch der Nederlanden et al., 2018; Vanden Bosch der Nederlanden, Hannon, & Snyder, 2015). Even though children are generally worse than adults at encoding and identifying individual sounds in busy scenes, all ages show a significant advantage for encoding speech that does not change with age, suggesting a robust and specific processing advantage for human speech sounds in mature language users. If processing biases for speech in complex scenes were evident during infants' first year, they could provide a potential mechanism for facilitating language learning.

Complex scene processing studies with adults and children have used the change deafness paradigm—the auditory analogue to change blindness (e.g., Simons, 2000)—to assess attentional biases towards real-world sounds (Vanden Bosch der Nederlanden et al, 2018; Vanden Bosch der Nederlanden, Hannon, & Snyder, 2015). In change deafness paradigms, listeners hear four different sounds played at the same time to create an auditory scene. After a period of silence, this 4-sound scene is played again and in about 50% of the trials, the second scene has one sound replaced by a new sound, while the other three sounds in the scene remain the same (see Gregg & Samuel, 2008; 2009). Adults' and children's detection of changing sounds in these complex auditory scenes is driven by the same two factors as their processing of faces in complex visual scenes: low-level salient physical features and high-level semantically meaningful combinations of features. Although the size of the acoustic change affects detection (see Method for how the size of acoustic change, called Euclidian distance, is calculated), mature listeners rely more on their semantic knowledge of real-world sounds to detect changes between scenes (Vanden Bosch der Nederlanden, Hannon, & Snyder, 2015). For example, adults are better at encoding and detecting changes to human vocal sounds, and speech in particular, compared to musical instruments, animal calls, and environmental sounds, regardless of the size of the acoustic change (Vanden Bosch der Nederlanden et al., 2018). Further, adults and children also have trouble detecting changes within the same sound category, such as when the sound of one dog barking changes to another dog barking, even when the acoustic distance is large (Gregg & Samuel, 2009; Vanden Bosch der Nederlanden, Snyder, & Hannon, 2016). Experienced listeners use their knowledge of real-world sound categories to preferentially attend to speech sounds in complex scenes, but it is unclear whether relatively inexperienced language learning infants would have similar processing biases for detecting changes to speech in complex scenes.

We examined whether infants have biases for processing speech in complex scenes. Specifically, we contrasted infants' detection of changes in speech sounds to their detection of changes in various non-speech sounds in complex auditory scenes. We compared speech to specific non-speech sound categories–musical instruments, animal calls, and water sounds–to test whether high-level specific sound categories or low-level physical properties like acoustic distance influenced infant processing. We measured the size of the acoustic change in a two-dimensional space by calculating the Euclidian distance between changing sounds within a trial (see Acoustic analyses in Method). For example, a change between a man speaking and a cello melody might be very similar in pitch (frequency) and would result in a small perceptual change. In contrast, changing from a woman speaking to a tuba melody would be a big change in pitch and timbre, which may be easier to detect.

In Experiment 1, we compared speech, musical instruments, and water sounds. Acoustic analyses showed that speech and musical instruments have similar acoustic features, allowing us to test whether, given equivalent acoustic distance, infants detect changes to speech better than other sounds. Water is more acoustically distant from both speech and musical instruments, allowing us to test how larger acoustic distances influenced infant detection. In Experiment 2, we further examined how sound categories and physical salience affect infant change detection by presenting speech and music to different infants and contrasting them with changes in animal sounds which have similar acoustic features, and water which has larger acoustic distance compared to the three other sounds. This study allowed us to compare the detection of music and speech changes in isolation from one another, while also examining the effect of physical salience and sound category on change detection in infancy. Across these two experiments, different non-speech sound categories allowed us to contrast whether infants' change detection was better predicted by physical salience (large acoustic changes), as for young infants' visual scene perception, or by a bias for speech, as for adults' and children's auditory scene perception.

We predicted that if infants' auditory biases take time to develop as in visual face processing studies, 5-month-olds may detect large acoustic changes (e.g., to water) better than changes to other sounds, including speech. At the same time, if infants already have biases for processing speech over other acoustically similar sound categories, they may preferentially detect changes to speech compared to music and animal calls (Vouloumanos & Werker, 2007). Better speech detection in complex scenes would suggest that infants' speech bias could help them isolate speech sounds and learn language in real-world auditory settings.

Method

Acoustic analyses

The magnitude of the acoustic change between the dropping and replacing auditory object was calculated for all change trials using pitch and harmonicity as the two dimensions of this Euclidian space (see Table 1). These dimensions were selected based on previous change deafness studies (Gregg & Samuel, 2009, Vanden Bosch der Nederlanden, Hannon, & Snyder, 2016) and based on previous studies of environmental sound similarity (Gygi, Kidd, & Watson, 2007). Speech and musical instrument sounds were nearly identical in terms of their average pitch and harmonicity, animal sounds were very similar in their harmonicity, and water sounds differed from most other sounds in pitch and harmonicity. The average Euclidian distance for each sound category is shown in Table 2. Euclidian distances were similar for speech, music, and animal change trials, but water change trials, because of their different acoustic features, resulted in a larger Euclidian distance than other sound categories. By including water sounds, we were able to look at how the size of the acoustic change influenced looking time.

Table 1. Acoustic characteristics (means) of each sound category: pitch, log of pitch, and harmonicity.

Stimuli	Pitch (Hz)	Log of Pitch	Harmonicity (dB)
Speech	162.92	2.19	14.63
Musical Instrument	153.87	2.19	13.42
Water	486.12	2.66	4.84
Animal	305.81	2.46	16.09

Table 2. Euclidian distances (mean of arbitrary units) measure the size of the acoustic change for each sound category in each experiment.

Change Type	Expt 1: Speech and Music	Expt 2: Speech Only	Expt 2: Music Only
Speech	7.37	5.92	–
Musical Instrument	6.76	–	5.65
Water	9.20	10.54	9.93
Animal	–	6.64	6.98

Experiment 1

Participants. Participants were 30 healthy, full term infants (13 females; *M* age: 5 months, 7 days; range 4 months, 15 days to 6 months, 1 day)[1]. Participants were recruited from maternity wards at local hospitals, and were reported as being healthy and having normal hearing at the time of testing. All parents reported that their children were learning English, and 19 parents reported additional languages for their infants including Spanish (6), French (3), Italian (3), Hebrew (2), Yiddish (1), Cantonese and Mandarin (1), Vietnamese (1), German (1), and Polish (1). An additional 13 infants (9 females) were excluded from analyses because of infant fussiness (3), experimenter error (5), looking at the screen for the maximum trial length on 5 or more trials (2), medical history (2), and parental interference (1). Parents gave informed consent on behalf of their infants and received a certificate and small toys or t-shirts as gifts for their participation. All procedures were approved by the IRB at [blinded for review] (IRB-FY2016-81, Divergent biases for conspecifics as early markers for autism spectrum disorders).

Design. To examine infant processing of complex auditory scenes, we modified the one-shot procedure used in change deafness studies (see Snyder, Gregg, Weintraub, & Alain, 2012) and adapted the alternating procedure common in visual

[1] Previous studies examining change deafness and different trial types have found η_p^2 ranging from .3 (Vanden Bosch der Nederlanden et al., 2018) to .79 (Vanden Bosch der Nederlanden et al., 2015). We estimated our sample size based on these effect sizes using a priori G*Power (Faul et al., 2009) estimated of within-subjects main effects of ANOVA, yielding 10 participants for 2 groups and 4 measurements. To be more conservative for infant testing results, we halved the effect size to .15 and this yielded a total sample size of 16 participants. We aimed to double this sample size, again because infant response data can be quite variable.

change blindness studies (e.g., Rensink, O'Regan, & Clark, 1997). This alternating paradigm switches between two static images with a brief blank image placed between them, with the length of time taken to detect the change indexing change blindness (longer reaction times indicate worse performance). This paradigm bears striking resemblance to infant looking time methods using stimulus-alternation preference procedures (Cowan, Suomi, & Morse, 1982; Best & Jones, 1998), that take advantage of infants' preference for novelty, with longer looking times to trials that alternate between two different visual displays or phonemes compared to looking at visual displays or phonemes from a single category. In our auditory paradigm, infants heard either alternating 2-sound scenes (see Figure 1), called change trials, or non-alternating scenes, called no change trials. Some trials involved a change in the speech sound, others involved changes in different non-speech sounds, e.g. music. Infants' looking time (compared with change trials) indexed whether or not they noticed the change between complex auditory scenes. Comparing changes for different sound categories allows us to examine whether infants' change detection is based on the size of the acoustic change (e.g., better for water) or on the sound category that is changing (better for speech, musical instruments).

Stimuli. Sounds were drawn from three categories: human speech, musical instruments, and water sounds. Two sounds from each category (woman's voice, man's voice; cello, tuba; bubbling water, draining water) were used to create 18 trials, 9 trials for each set with one sound from each category in each set (Set 1: woman's voice, cello, bubbling water; Set 2: man's voice, tuba, draining water). Half the infants were tested with Set 1 and half the infants were tested with Set 2. Trials were composed of 2-object scenes. A 2-object “scene” consisted of 2 overlapping sounds 1000 ms in length with simultaneous onsets and offsets. These scenes were concatenated with 500 ms silent inter-stimulus intervals to make 43-second long sound trials. In a no change trial, the same 2-object scene was repeated for the entire length of the trial. In a change trial, the 2-object scene alternated with a different 2-object scene in which one of the sounds was the same and one of the sounds changed. For example, in a speech change trial, the woman’s voice + draining water scene would alternate with the cello + draining water scene (see Figure 1A-C). Each infant heard 9 trials in total: 3 no change trials and 6 change trials: 2 speech change trials, 2 music change trials, and 2 water change trials (see Appendix for specific sound combinations).

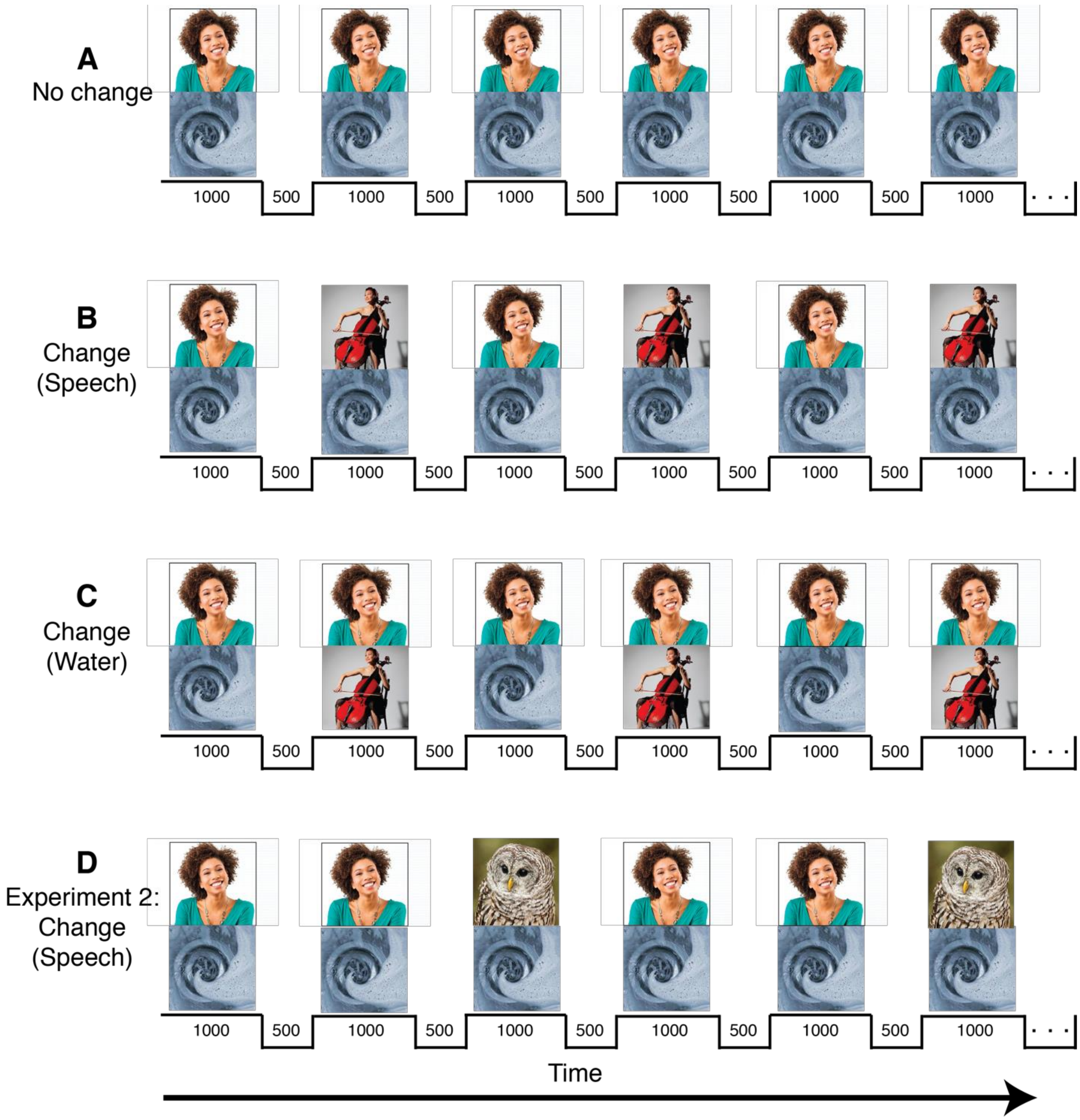

Figure 1. Example of no change and change trials for Experiment 1 (A-C) and Experiment 2 (A and D), which had a same-same-different structure instead of simple alternations between scenes.

Procedure. Infants were tested in a sound-attenuated room using an infant-controlled sequential preferential looking procedure (e.g., Cooper & Aslin, 1994, Vouloumanos & Werker, 2004), run in Habit 2.1.25 (Oakes, Sperka, & Cantrell, 2015). In this procedure, infants controlled the onset and offset of each trial by looking at or away from a central monitor. Infants sat on a parent's lap 35" (89 cm) in front of a 30" (76.25 cm) computer monitor. Parents wore headphones playing a sound medley to mask the experimental sounds. At the start of the experiment, infants' attention was drawn to the monitor by a colorful expanding and contracting circle. Once infants fixated on the monitor, a stationary black and white checkerboard appeared in tandem with one set of sounds presented at a mean amplitude of 60 dB (±5 dB). An experimenter, blind to the stimulus condition, recorded infants' fixations to and away from the monitor during the study. Trials ended if infants looked away for longer than 2 seconds or if the maximum trial length was reached (43 seconds). A new trial began once infants looked at the expanding and contracting circle on the center monitor.

Infant looking time was coded online and verified offline by coding frame-by-frame looking to the monitor using 30 frame-per-second videos. The more precise offline looking times were used in the analyses. Average looking times for each of the four trial types were calculated for each participant: no change, speech change, music change, water change. Trials with missing looking time values were not include in the final looking time average for each participant. All figures are reported using within-subjects error bars given the repeated-measures design of our study (Cousineau, 2005).

Results and Discussion

Looking times for the sound dropping out of the first scene[2] were submitted to a repeated measures Analysis of Variance (ANOVA) with no change, speech change, music change, and water change trials as levels of the within-subjects factor sound and stimulus set (set 1, set 2) as a between-subject factor. There was no main effect (p = .623) or interaction (p = .459) with stimulus set. There was a main effect of trial type $F(3,84) = 4.425$, $p = .013$, $\eta_p^2 = .136$. Planned comparisons first examined whether infants detected the three sound category changes relative to no change trials: Infants looked longer at speech change (p =. 019) and music change trials (p = .005) than no change trials, but water change trials did not differ from no change trials (p = .355). Three planned comparisons compared the change conditions to each other to test our main question of whether semantic category or acoustic distance influenced infant change detection. Music change and speech change trials did not differ from one another (p = .102, see Figure 2). Water change trials did not differ from speech change trials (p = .376), but looking time was significantly higher for music than for water change trials (p = .041).

[2] Each change trial inherently involves a sound dropping out of one scene and a sound replacing it in the next scene. As such, change trials can be categorized by either the dropping or replacing sound. For instance, in Figure 1B, speech drops out of the first scene and is replaced by the cello. This change trial could be categorized as either a speech (dropping sound) or musical instrument (replacing sound) change trial. All analyses were completed based on both dropping and replacing sound groupings. Replacing sounds resulted in only a marginal main effect of trial type, $F(3,84) = 2.682$, $p = .067$, $\eta_p^2 = .087$, whereas the effect size for dropping sounds was larger and is reported in the main text.

Together these results suggest that infants looked longer at music and speech changes than no change trials. That is, when infants listened to multiple sounds presented simultaneously in a scene, they noticed changes to musical instruments and human speech equally well. At the same time, detection of changes in musical instruments, but not in human speech, was better than detection of changes in water sounds suggesting some attentional priorities for musical instruments in auditory scene processing. Attentional priorities may be important for guiding infants' and children's attention toward relevant sounds in our environment and for scaffolding learning in the real world (e.g., Frank, Vul, & Johnson, 2009; Scerif, 2010).

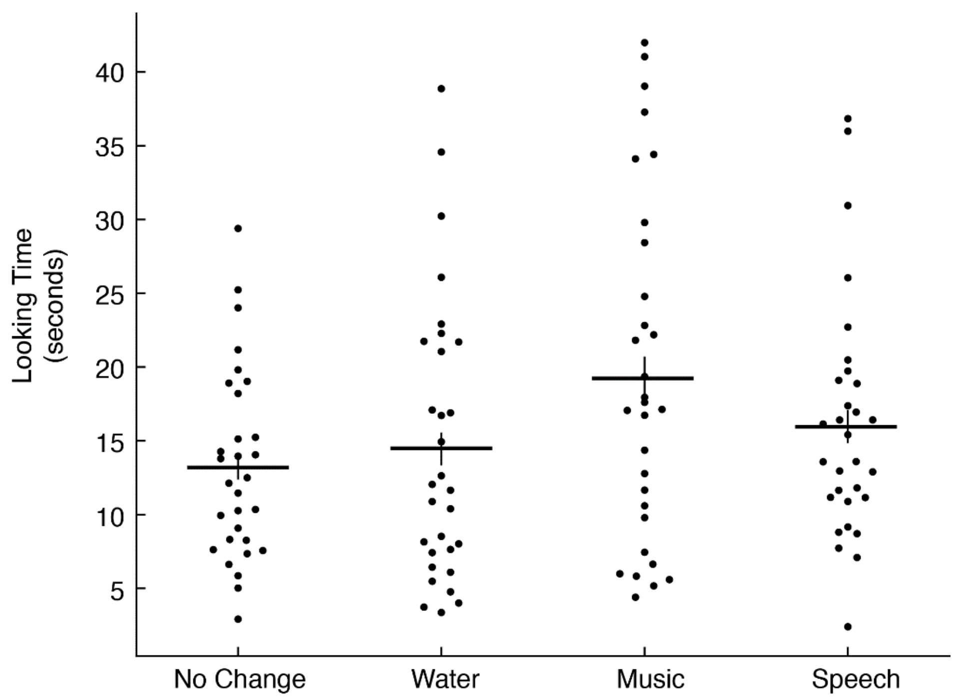

Figure 2. Infants looked longer at music and speech change trials than no change trials; water change trials were not different from no change trials. Speech change trials also did not differ from water change trials. Error bars are within-subjects standard error.

However, these results are not entirely straight-forward as speech change trials did not differ significantly from water change trials and water change trials did not differ from no change trials. The design in of Experiment 1 made it difficult to tease apart looking time for speech and music separately as they were often heard together and, in some trials, music was heard when speech changed and vice-versa. Further, the alternating design suggested two possibilities for how to analyze change trials, grouping results either by the sound dropping or replacing in each scene (see Footnote 1).

Experiment 2

The results of Experiment 1 suggest that music and speech were both salient changes to infants, but the interpretation of this effect is tempered by the lack of a difference between speech and water change trials, and in turn, the lack of difference between water and no change trials. To further examine change detection of different sound categories, in Experiment 2, we separated music and speech into different conditions and adopted a same-same-different alternating design. The same-same-different alternating paradigm has been used in previous visual change deafness studies and more clearly highlights the stable sound as it occurs twice in every triplet. It is important to note that speech and musical instruments are well-matched in their acoustic features and so differences across conditions would not be due to pitch, harmonicity, or the size of the acoustic change. We compared speech and music to animal sounds because their acoustic features and the size of the acoustic change are comparable (see Table 1).

Participants. Participants were 36 healthy, full term infants (10 females; *M* age: 5 months, 12 days, range 4 months, 15 days to 6 months, 0 days).[3] None of the participants had a history of ear infections or had a cold/ear infection at the time of the study. Parents reported that all children were learning English, and 29 parents reported additional languages for their infants including with Spanish (17), Hebrew (2), Russian (2), Italian (2), German (2), Farsi (1), Fulani (1), Tigrinya (1), and Hindi (1). An additional 6 infants were excluded from the current analyses because of equipment failure such that they did not hear any sound during the study (4), or because of infant fussiness (2). Infants were randomly assigned to either the music condition (17 infants, 4 females, mean age 5 months, 12 days) or the speech condition (19 babies, 6 females, mean age 5 months, 12 days). Parents gave informed consent on behalf of their infants and received a certificate and small toys or t-shirts as gifts for their participation. All procedures were approved by the [blinded for review] (IRB-FY2016-81, Divergent biases for conspecifics as early markers for autism spectrum disorders).

Stimuli. Sounds were drawn from four sound categories: human speech, musical instruments, water sounds, and animal sounds. Speech and music were presented in separate conditions. For speech, two sounds of each type (woman’s voice, man’s voice; bubbling water, draining water; owl hoot, sheep bah) were used to create 18 trials, 9 trials for each set, with one sound from each category in each set (Set 1: man’s voice, owl hoot, draining water; Set 2: woman’s voice, sheep bah, bubbling water). For music, two sounds of each type (cello, tuba; bubbling water, draining water; sheep bah, owl

[3] Sample sizes were collected based off of the η_p^2 obtained from the Experiment 1 η_p^2=.136, using G*Power (Faul et al., 2009) to calculate that a total of 18 participants were needed per condition.

hoot) were used to create 18 trials, 9 trials for each set with one sound from each category in each set (Set 1: cello, owl hoot, draining water; Set 2: tuba, sheep bah, bubbling water). About half the infants were tested with Set 1 and the others were tested with Set 2. All trials were composed of 2-object scenes created in the same manner as Experiment 1, with 1000 ms overlapping sounds with simultaneous onsets and offsets with 500-ms silent intervals. Unlike Experiment 1, change scenes had a same-same-different 2-object scene structure and 10 repetitions of that sequence. No change trials consisted of a same-same-same sequence (Figure 1A). All trials were 45 seconds in length. For example, in a speech change trial, the woman's voice + draining water scene would be presented twice and followed by the owl + draining water scene (see Figure 1D). Each infant heard 9 trials in total: 3 no change trials and 6 change trials: 2 speech or music change trials, 2 water change trials, and 2 animal change trials.

Procedure. Identical to Experiment 1. Infants were pseudo-randomly assigned to the music or speech conditions. About half of the infants in each study were assigned to one of two stimulus sets (Speech: Set 1 = 9 infants, Set 2 = 10; Music: Set 1 = 8, Set 2 = 9 infants).

Results and Discussion

To confirm that infants were equally attentive during both sound conditions, total looking time across all trial types was compared between the speech and music conditions in a one-way ANOVA with condition as a between-subjects factor. There was no difference between total looking time for the music (55.82 s) and speech (60.53 s) conditions, $F(1, 34) = .406$, $p = .528$, $\eta_p^2 = .012$.

We analyzed whether looking time to change trials differed as a result of sound category and/or acoustic salience in two separate ANOVAs for the speech and music conditions. Looking time was analyzed with trial type (no change, water change, animal change, speech or music change) as the within-subjects variable and stimulus set (Set 1, Set 2) as a between subject variable.

Speech. There was a main effect of trial type $F(3,51) = 3.990$, $p = .024$, $\eta_p^2 = .190$, but no main effect for stimulus set, $F(1,17) = .596$, $p = .451$, $\eta_p^2 = .034$, or interaction, $F(3,51) = .137$, $p = .888$, $\eta_p^2 = .008$. Planned comparisons first examined whether infants detected the three sound changes relative to no change trials: speech change ($p = .005$) and water change trials ($p = .033$) differed from no change trials, but animal change trials did not ($p = .430$). Three planned comparisons examined how looking time for change trials differed from each other to again examine how sound category and acoustic change magnitude influenced looking time. Speech change trials elicited longer looking times than animal change trials ($p = .045$), but did not differ from water change trials ($p = .614$) and animal change trials did not differ from water change trials ($p = .132$, see Figure 3). Thus, infants looked longer at speech trials than no change and animal change trials, but they looked equally at speech and water change trials.

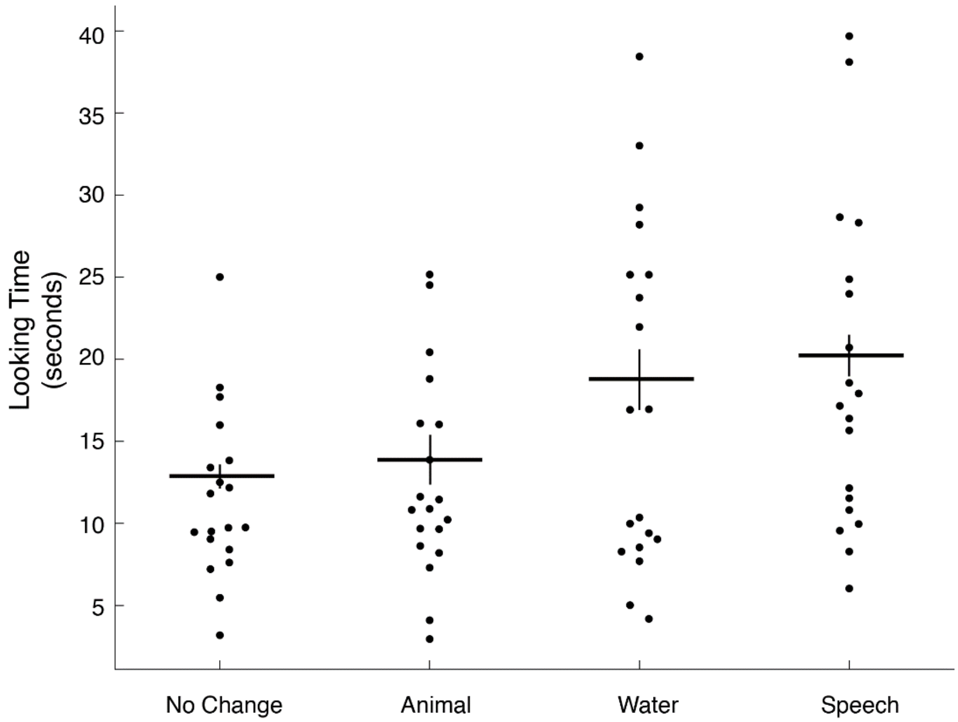

Figure 3. Mean looking times to no change, animal, water, and speech trials in Experiment 2. Error bars are within-subjects standard error.

Music. There were no main effects of trial type, $F(3,45) = 1.651$, $p = .191$, $\eta_p^2 = .099$ or set, $F(1,15) = 1.250$, $p = .281$, $\eta_p^2 = .077$, and no interaction ($p = .155$). These findings suggest that there was no difference between music change trials and no change trials or any other change trials in Experiment 2 as is clear from Figure 4.

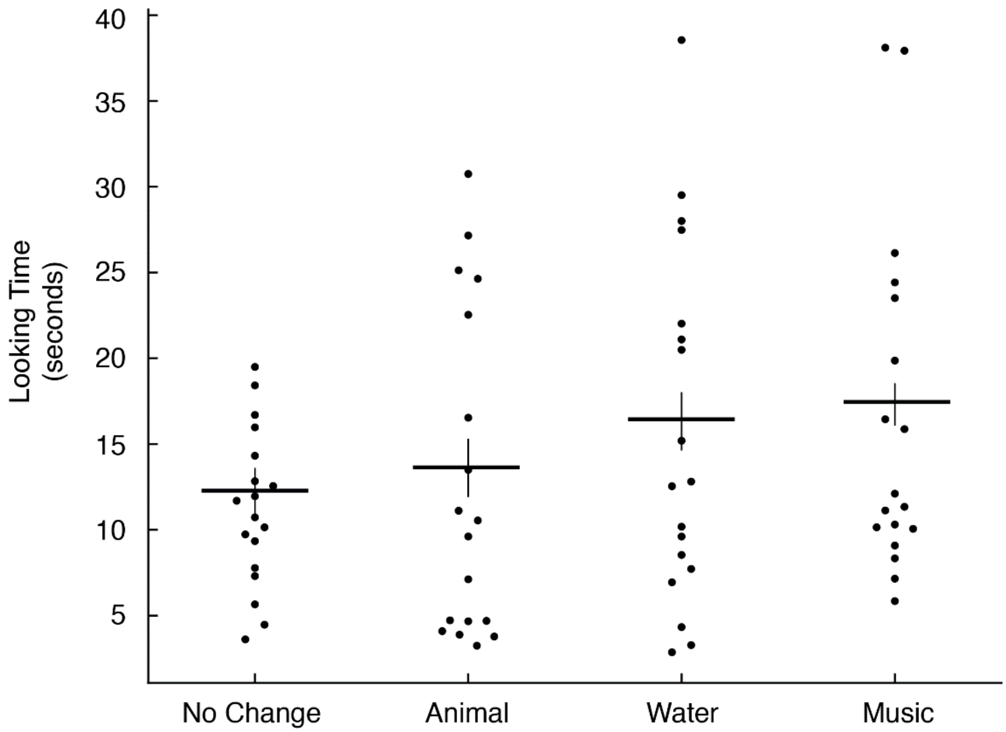

Figure 4. Mean looking times to no change, animal, water, and music trials in Experiment 2. Error bars are within-subjects standard error.

Together these findings suggest that infants attend to changes to speech sounds more than changes to music sounds in complex scenes composed of real-world sounds, despite both sound categories having similar acoustic properties. As in Experiment 1, speech change trials did not differ from water change trials. We further analyzed whether infants' looking time was influenced by the magnitude of the change for different sound types.

Euclidian distance. One possible explanation for the longer than expected looking times to water sounds may be the Euclidian distance – that is, the size of the acoustic change within a change trial. The average size of the change was larger for water change trials compared to the other change types (see Table 1 and Table 2). To test this explanation, we examined whether infant looking times correlated with

Euclidian distance. There was a significant correlation between Euclidian distance and participants' looking time for the speech condition, $r(18) = .585$, $p = .011$, but not the music condition, $r(18) = .367$, $p = .135$. This suggests that infants were sensitive to the size of the acoustic change in the speech condition, but not the music condition. Infants' looking time did not correlate with Euclidian distance for the combined speech and music scenes from Experiment 1, $r(18) = .338$, $p = .170$. Looking time may only have correlated with the speech condition because this had the largest spread in Euclidian distances compared to the other two conditions (see Table 2), suggesting that infants may be sensitive to the size of the change, but require a larger spread in acoustic change magnitude to show this sensitivity.

The findings from Experiment 2 suggest that infants are biased toward listening to human speech in complex scenes, but they are also sensitive to acoustic salience, as they detected water changes in both conditions. Infants do not appear to show any such bias for music changes. Infants' equal detection of music changes and speech changes in Experiment 1 may have been due to speech also being played during the music change trials which heightened infants' interest in these scenes. By presenting speech and music sounds separately, we could examine the independent contribution of these different sound categories to differences in looking time in Experiment 2.

General Discussion

How do infants learn language in the real world when, so often, many different sounds are competing for the infants' attention? We examined whether infants' early emerging listening biases for speech may be evident in scene processing, giving them a tool to detect and process speech from noisy environments. In two experiments, 5-month-olds looked longer when speech sounds changed across auditory scenes compared to when acoustically comparable non-speech sounds changed. In Experiment 1, infants looked longer for changes in music and speech compared to no change trials. In Experiment 2, when infants heard either speech changes or music changes in separate conditions, infants looked longer for speech changes, but not for musical instrument changes compared to no change trials. This suggests that the longer looking times for musical instruments and speech sounds observed in Experiment 1 may have been due to both sounds being presented in the same scene. At the same time, in both experiments, infants looked equally at changes in speech and at changes in water sounds, suggesting that acoustic distance also influences how infants detect changes in sounds. This study provides the first evidence that infants preferentially process changes to speech in complex acoustic scenes with multiple overlapping sounds, as well as being sensitive to low-level physical properties such as acoustic distance.

Given comparable acoustic distance between sounds within sound categories, infants looked longer to speech changes than changes in music, and animal sounds. This is consistent with previous work with children and adults showing that listeners of all ages relied more heavily on their knowledge of sound categories than on the size of an acoustic change to listen to the sounds around them (Vanden Bosch der Nederlanden, Hannon, Snyder, 2016). Like adults and 6-10-year-old children, 5-month-olds appear to use auditory listening strategies that are more mature, biasing attention toward changes in semantically meaningful categories, even when the acoustic changes were small. Listening preferentially to the sounds that are most communicatively relevant can make the task of listening to multiple sounds within a scene much more efficient. Instead of having to encode and store changes to a number of different sounds in working memory (Snyder, Gregg, Weintraub, & Alain, 2012), filtering out sounds that are less relevant for communication and biasing attention towards sounds with high communicative value could be an important step for bootstrapping language learning in the first months of life. As such, detecting small changes in speech sounds may be an important pre-requisite for speech perception and language learning in the real world. Future work could examine whether a stronger bias for detecting changes in the human voice and speech in complex scenes is related to phonological awareness or word learning by the first year of life.

Beyond their sensitivity to specific sound categories, infants were sensitive to the size of the acoustic change. This was evident when infants looked longer to water sounds changing compared to no change trials, and in their equal looking to changes in water and changes in speech in both experiments. Moreover, looking times were positively correlated with Euclidian distance across participants in Experiment 2 for the speech condition. These results directly align with previous change deafness work suggesting that children and adults use both their knowledge of sound categories and the magnitude of an acoustic change to detect changes across auditory scenes (Gregg & Samuel, 2009; Vanden Bosch der Nederlanden, Hannon, & Snyder, 2016). By 5 months of age, perceptual abilities to detect changes based on their stimulus features are already in place. These findings somewhat mirror the development of change detection in complex visual scenes. Infants looked longer to salient features in childfriendly movies, which declined between 3- and 9-months of age, while looking to faces increased over this period (Frank, Vul & Johnson, 2009). Although infants in the current study already show a bias toward human speech over large acoustic changes, further studies should examine whether infants preferentially attend to large, salient changes early in development across modalities and only begin to attend preferentially to sound categories like the human voice and speech with attentional maturity (Frank, Amso, & Johnson, 2014) or with experience.

One limitation of the current study is the small number of simultaneously presented sounds in each scene. As this was the first study to examine change detection in the auditory modality in infancy, we limited scenes to two overlapping sounds. Previous change deafness literature found that children and adults are biased toward the human voice with 4-sound scenes (Vanden Bosch der Nederlanden, Snyder, & Hannon, 2016). Other change deafness studies have examined change detection in scene sizes up to 8 simultaneously presented sounds (Eramudugolla et al., 2005; Gregg, Irsik, & Snyder, 2017). As the young language learner's acoustic world could contain a large number of sounds in a typical scene, future studies will need to characterize how many sounds infants can hear in a given scene while still detecting changes between scenes, and whether the bias for speech persists with more simultaneous sounds.

Our findings show that young infants building their language system not only have robust and early emerging preferences for listening to speech over other sounds (e.g., Vouloumanos & Werker, 2007; 2004), but they also detect changes in speech sounds better than changes of a similar magnitude for other sounds–over and above the size of acoustic changes–when sounds are presented simultaneously in complex auditory scenes characteristic of the real world.

References

Best, C. T. & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behaviour and Development, 21*, 295. *doi:* **10.1016/S0163-6383(98)91508-9**

Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. Cambridge, MA, US: The MIT Press.

Cooper, R. P. & Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development, 65*(6), 1663-1677. *doi:* **10.1111/j.1467-8624.1994.tb00841.x**

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42-45. *doi:* **10.20982/tqmp.01.1.p042**

Cowan, N., Suomi, K., & Morse, P. A. (1982). Echoic storage in infant perception. *Child Development, 53*(4), 984-990. *doi:* **10.2307/1129138**

Eramudugolla, R., Irvine, D. R., McAnally, K. I., Martin, R. L., & Mattingley, J. B. (2005). Directed attention eliminates 'change deafness' in complex auditory scenes. *Current Biology, 15*(12), 1108-1113. *doi:* **10.1016/j.cub.2005.05.051**

Farroni T, Johnson MH, Menon E, Zulian L, Faraguna D, Csibra G. (2005). Newborns' preference for face- relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences (USA), 102*,17245–17250.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Fletcher-Watson, S., Leekam, S. R., Benson, V., Frank, M. C., & Findlay, J. M. (2009). Eye-movements reveal attention to social information in autism spectrum disorder. *Neuropsychologia, 47*, 248-257. *doi:* **10.1016/j.neuropsychologia.2008.07.016**

Frank, M. C., Amso, D., & Johnson, S. P. (2014). Visual search and attention to faces during early infancy. *Journal of Experimental Child Psychology, 118,* 13-26. *doi:* **10.1016/j.jecp.2013.08.012**

Frank, M. C., Vul, E. & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition, 110*(2), 160-170. *doi:* **10.1016/j.cognition.2008.11.010**

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy, 14*(5), 550-562. *doi*: **10.1080/1525oooO9O3 144199**

Gregg, M. K., Irsik, V. C., & Snyder, J. S. (2017). Effects of capacity limits, memory loss, and sound type in change deafness. *Attention, Perception & Psychophysics, 79*(8), 2564-2575. *doi:* **10.3758/s13414-017-1416-4.**

Gregg, M. K. & Samuel, A. G. (2008). Change deafness and the organizational properties of sounds. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 974-991. *doi:* **10.1037/0096-1523.34.4.974**

Gregg, M. K. & Samuel, A. G. (2009). The importance of semantics in auditory representations. *Attention, Perception & Psychophysics, 71*(3), 607-619. *doi:* **10.3758/APP.71.3.607**

Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception & Psychophysics, 6*(69), 839-855. *doi:* **10.3758/BF03193921**

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*(3-4), 159-207. *doi:* **10.1006/cogp.1999.0716**

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

Lim, S. J., Lacerda, F., & Holt, L. L. (2015). Discovering functional units in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance, 41*(4), 1139-1152. *doi:* **10.1037/xhp0000067**

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics, 66*(8), 1268-1281. *doi:* **10.3758/BF03194997**

Newman, R. S. & Jusczyk, P. W. (1996). The cocktail party effect in infants. *Perception & Psychophysics, 58*(8), 1145-1156. *doi:* **10.3758/BF03207548**

Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., Dow, K. A. (2006). Infants' early ability to segment the conversation speech signal predicts later language development. A retrospective analysis. *Developmental Psychology, 42*(4), 643-655. *doi*: **10.1037/0012-1649.42.4.643**

Oakes, L.M., Sperka, D. J., & Cantrell, L. (2015). Habit 2. *Unpublished software. Center for Mind and Brain, University of California, Davis.*

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368-373. *doi:* **10.1111/j.1467-9280.1997.tb00427.x**

Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science, 12*(1), 94-99. *doi:* **10.1111/1467-9280.00317**

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294): 1926-1928. *doi:* **10.1126/science.274.5294.1926**

Scerif, G. (2010). Attention trajectories, mechanisms and outcomes: at the interface between developing cognition and environment. *Developmental Science, 13*(6), 805-812. *doi:* **10.1111/j.1467-7687.2010.01013.x**

Shultz, S. & Vouloumanos, A. (2010). Three-month-olds prefer speech to other naturally occurring signals. *Language Learning and Development, 6*, 241-257. *doi:* **10.1080/15475440903507830**

Simion F, Macchi Cassia V, Turati C, Valenza E. (2001). The origins of face perception: specific versus non- specific mechanisms. *Infant and Child Development,10*,59–65.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7*(1-3), 1-15. *doi:* **10.1080/135062800394658**

Simons, D. J. & Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends in Cognitive Sciences, 9*(1), 16-20. *doi:* **10.1016/j.tics.2004.11.006**

Snyder, J. S., Gregg, M. K., Weintraub, D. M., & Alain, C. (2012). Attention, awareness, and the perception of auditory scenes. *Frontiers in Psychology, 3,* Article ID 15. *doi:* **10.3389/fpsyg.2012.00015**

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language, 60*(2), 252-269. *doi:* **10.1016/j.jml.2008.11.003**

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7*(1), 53-71. *doi:* **10.1207/s15327078in0701_5**

Vanden Bosch der Nederlanden, C. M., Snyder, J. S., & Hannon, E. E. (2015). Everyday musical experience is sufficient to perceive the speech-to-song illusion. *Journal of Experimental Psychology: General, 144*(2), e43-e49. *doi:* **10.1037/xge0000056**

Vanden Bosch der Nederlanden, C. M., Snyder, J. S., & Hannon, E. E. (2016). Children use object-level category knowledge to detect changes in complex auditory scenes. *Developmental Psychology, 52*(11), 1867-1877. *doi:* **10.1037/dev0000211**

Vanden Bosch der Nederlanden, C. M., Zaragoza, C., Rubio-Garcia, A., Clarkson, E., & Snyder, J. S. (2018). Change detection in complex auditory scenes is predicted by auditory memory, pitch perception, and years of musical training. *Psychological Research,* 1-17. *doi:* **10.1007/s00426-018-1072-x**

Vouloumanos, A. & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science, 7*(3), 270-276. *doi:* **10.1111/j.1467-7687.2004.00345.x**

Vouloumanos, A. & Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science, 10*(2), 159-164. *doi:* **10.1111/j.1467-7687.2007.00549.x**

Vouloumanos, A. & Waxman, S. R. (2014). Listen up! Speech is for thinking during infancy. *Trends in Cognitive Neuroscience, 18*(12), 642-646. *doi:* **10.1016/j.tics.2014.10.001**

Waxman, S. R. & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*(3), 257-302. *doi:* ***10.1006/cogp.1995.1016***

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology, 34*(6), 1289-1309. *doi:* **10.1037/0012-1649.34.6.1289**

Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7*(1), 49-63. *doi:* **10.1016/S0163-6383(84)80022-3**

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science, 21*(4), 221-226. *doi:* **10.1177/0963721412449459**

Appendix

Set 1				Set 2			
Trial	Scene 1	Scene 2	Sound Change Category	Trial	Scene 1	Scene 2	Sound Change Category
1	Female Speaking Bubbling Water	Female Speaking Bubbling Water	No Change	10	Male Speaking Draining Water	Male Speaking Draining Water	No Change
2	Female Speaking Bubbling Water	Cello Bubbling Water	Speech Change	11	Male Speaking Draining Water	Tuba Draining Water	Speech Change
3	Female Speaking Bubbling Water	Female Speaking Cello	Water Change	12	Male Speaking Draining Water	Male Speaking Tuba	Water Change
4	Female Speaking Cello	Female Speaking Cello	No Change	13	Male Speaking Tuba	Male Speaking Tuba	No Change
5	Female Speaking Cello	Bubbling Water Cello	Speech Change	14	Male Speaking Tuba	Draining Water Tuba	Speech Change
6	Female Speaking Cello	Female Speaking Bubbling Water	Music Change	15	Male Speaking Tuba	Male Speaking Draining Water	Music Change
7	Bubbling Water Cello	Bubbling Water Cello	Water Change	16	Draining Water Tuba	Draining Water Tuba	Water Change
8	Bubbling Water Cello	Female Speaking Cello	Water Change	17	Draining Water Tuba	Male Speaking Tuba	Water Change
9	Bubbling Water Cello	Bubbling Water Female Speaking	Music Change	18	Draining Water Tuba	Draining Water Male Speaking	Music Change