

© Copyright 2018

Monika-Maria Oster

Infants' Use of Temporal Cues in the Segregation of Concurrent Sounds

Monika-Maria Oster

A dissertation

submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Lynne A. Werner, Chair

Adrian KC Lee

Lori Leibold

Program Authorized to Offer Degree:

Speech and Hearing Science

University of Washington

**Abstract**

Infants' Use of Temporal Cues in the Segregation of Concurrent Sounds

Monika-Maria Oster

Chair of the Supervisory Committee:  
Lynne A. Werner, Professor Emerita  
Speech and Hearing Sciences

Infants have greater difficulties processing speech in the presence of competing sounds than adults. The mature auditory system solves this task in part by separating sounds that have different acoustic characteristics such as onset, temporal envelope or fundamental frequency. The current work evaluated whether 3- and 7-month-old infants use two different acoustic characteristics, onset asynchrony and temporal envelope, similarly to adults in the segregation of concurrent vowels. To this end, listeners were presented with superimposed vowels, consisting of 2 concurrent vowels spoken by two different-sex talkers, and were trained to respond to one specific target vowel. Three studies were conducted to evaluate infants' use of 1) onset asynchrony, 2) envelope differences, and 3) combined onset asynchrony and envelope differences. The results indicated that 3-month-old infants were better able to segregate vowels with different onset, but not with different envelopes or a combination of onset and envelope

differences. In contrast, by 7-months of age, infants used isolated and combined differences in onset and envelopes to the same extent as adults. Thus, after 7-months of age, an inability to use differences in onset and temporal envelopes cannot account for infants' difficulties processing competing speech sounds.

# TABLE OF CONTENTS

List of Figures .....	iii
List of Tables .....	iv
Chapter 1 Introduction .....	1
Chapter 2 Infants' use of onset asynchrony cues.....	4
I. INTRODUCTION .....	4
II. EXPERIMENT 1 .....	8
A. Method .....	8
B. Results .....	14
C. Discussion .....	17
III. EXPERIMENT 2 .....	18
A. Methods.....	19
B. Results .....	21
C. Discussion .....	22
IV. GENERAL DISCUSSION .....	23
Chapter 3 Infants' use of Envelope cues.....	26
I. INTRODUCTION .....	26
II. EXPERIMENT 1 .....	29
A. Method .....	29
B. Results .....	36
C. Discussion .....	39

III. EXPERIMENT 2 .....	40
A. Method .....	41
B. Results .....	42
C. Discussion .....	44
IV. GENERAL DISCUSSION .....	45
Chapter 4 Infants' use of combined onset asynchrony and envelope cues.....	48
I. INTRODUCTION.....	48
II. EXPERIMENT 1 .....	50
A. Method .....	50
B. Results .....	57
C. Discussion .....	62
III. EXPERIMENT 2 .....	63
A. Method .....	64
B. Results .....	65
C. Discussion .....	67
IV. SUMMARY AND CONCLUSIONS.....	68
Chapter 5 Conclusions and Future Directions .....	70
Bibliography .....	73

## LIST OF FIGURES

Figure 2-1: F1/F2 space for the vowels used in experiments 1 and 2.....	10
Figure 2-2 Individual $d'$ scores for infants and adults.....	15
Figure 2-3 Average $d'$ in each condition by age group.....	16
Figure 3-1 schematic representation of double-vowel pairs.....	32
Figure 3-2 Individual $d'$ scores in the baseline and envelope cue condition.....	37
Figure 3-3 Average $d'$ in the baseline and envelope cue conditions.....	38
Figure 3-4 Box-and-whisker plot of trials-to-criterion as a function of cue condition.....	44
Figure 4-1 schematic representation of vowel pairs in the combined cue condition.....	53
Figure 4-2 Individual $d'$ scores in the baseline and combined cue conditions.....	58
Figure 4-3 Average $d'$ in the baseline and the combined cue conditions.....	59
Figure 4-4 Average $d'$ for infants and adults in three experiments.....	61
Figure 4-5 Box-and-whisker plot of trials to criterion with and without combined cues.....	66

## LIST OF TABLES

Table 2-1 Results of mixed models GEE statistical analysis of $d'$ .....	17
Table 3-1 1 <sup>st</sup> and 2 <sup>nd</sup> formant (F1, F2) of all vowels from each talker. ....	31
Table 3-2 Results of mixed models analysis: the effect of age and envelope cue on $d'$ ...	39
Table 4-1 Results of mixed models analysis: the effect of age and combined cues on $d'$	60
Table 4-2 Results of mixed models analysis: the effect of age, presence and type of cue on $d'$ . .....	61

## **ACKNOWLEDGEMENTS**

First and foremost I would like to thank my incredible mentors, who have provided me with the guidance, encouragement, constructive criticism, and inspiration that helped me complete this journey. Additionally, I would like to thank the members of the Infant Hearing Lab, the families and individuals who participated in this research, and my funding sources. Without them, this work would not have been possible. Lastly, I am deeply grateful to my family and friends for their unwavering and loving support.

## CHAPTER 1 INTRODUCTION

It is well documented that separating speech from competing sounds presents greater difficulties for infants than for adults (Leibold, Bonino, & Buss, 2016; Newman, 2005; Newman & Morini, 2017; Nozza, Rossman, Bond, & Miller, 1990; Oster & Werner, 2017; Werner, 2013). The reasons for infants' difficulties are not well understood. Processing competing sounds is a complex problem since the frequency components from all the sound sources in the environment arrive at the ear as a sound mixture. However, frequency components that originate from the same source have similar acoustic characteristics, such as fundamental frequency, temporal envelope or onset. Research in adults has shown that successfully separating competing sounds relies on at least three processes: 1) accurate encoding of sound components, 2) grouping components that have similar acoustic characteristics and segregating those that differ, and 3) auditory attention. It appears clear that the ability to encode auditory information accurately cannot account for infants' difficulties, since spectral and temporal encoding is quite mature early in infancy (Buss, Hall, & Grose, 2012). For simple stimuli, where infant-adult differences are relatively small, nonsensory factors such as listening strategy or inattention can largely account for age-related differences (Bargones, Werner, & Marean, 1995; Werner & Bargones, 1991; Werner & Boike, 2001). In contrast, the large infant-adult differences in performance with complex stimuli such as detecting a tone amidst variable frequency components (e.g., Leibold & Werner, 2006) or processing speech in competing speech (e.g., Leibold, Taylor, Hillock-Dunn, & Buss, 2013), cannot fully be explained by those factors. It is possible that in these cases, age-related differences are due to infants' (in-)ability to use acoustic characteristics, also known as auditory cues, to effectively group and segregate the frequency components of competing sounds (Leibold, 2012).

Despite a number of studies that have evaluated infants' auditory cue use, a cohesive account of its development has yet to be established. Prior investigations have primarily focused on spectral cues, have not systematically varied cues, have used methods that preclude the inclusion of different age groups, and have almost exclusively used non-speech stimuli (e.g., Bendixen, Haden, Nemeth, Farkas, Torok, & Winkler, 2015; Demany, 1982; Folland, Butler, Payne, & Trainor, 2015; Folland, Butler, Smith, & Trainor, 2012; Winkler, Kushnerenko, Horváth, Čeponienė, Fellman, Huotilainen, Näätänen, & Sussman, 2003).

The goal of the current work was to expand knowledge on auditory cue use development in infancy and on how it might contribute to the age-related differences observed in the processing of competing sounds. To this end, infants' ability to use two temporal cues, onset asynchrony and envelope cues, in the segregation of concurrent speech sounds was evaluated. The studies reported here used a single behavioral method, the observer-based procedure, which allows the inclusion and direct comparison of infants and adults (Werner, 1995). In the experimental task, listeners were presented with two different concurrent vowels, spoken by a male and a female talker. Listeners were trained to respond to one specific target vowel, either male /u:/ or male /i:/. The ability to detect this target vowel was compared between conditions in which the competing vowels had either similar or different temporal cues. For 3-month-old infants an increase in the proportion of infants who learned the task with a temporal cue compared to the proportion who learned the task without was taken as evidence that the cue supported vowel segregation. For 7-month-old infants and adults an increase in performance, measured in  $d'$ , with a temporal cue over performance without a temporal cue was taken as evidence that the cue supported vowel segregation. Three separate studies, each addressing a different set of research questions, were conducted.

The first study investigated whether 3- and 7-month-old infants use onset asynchronies to segregate competing vowels and whether cue use improved with age and onset asynchrony duration. Since onset asynchrony is a reliable cue for adults (e.g., Hall, Buss, & Grose, 2006; Hedrick & Madix, 2009; Lentz & Marsh, 2006; Summerfield, Culling, & Fourcin, 1992; Valentine & Lentz, 2012) and is available to infants (Bendixen et al., 2015), it was expected that all listeners would gain a benefit from this cue. An increase in onset asynchrony was expected to improve performance (Lentz & Marsh, 2006), but more for infants than for adults given that children tend to require larger acoustic cues to achieve adultlike segregation (Sussman, Wong, Horvath, Winkler, & Wang, 2007).

The second study investigated whether 3- and 7-month-old infants use envelope cues in concurrent vowel segregation, and whether 7-month-olds use them to the same extent as adults. Adults were expected to gain a benefit from differences in amplitude envelopes (Moore & Gockel, 2012). In contrast, infants were expected to gain little benefit from envelope cues, given their insensitivity to amplitude modulation (AM) (Horn, Hutton Gerhards, & Werner, 2016; Werner, 2006; Werner, Folsom, Mancl, & Syapin, 2001) and the finding that children seem to gain little benefit from envelope cues (Hall, Buss, & Grose, 2008).

The last study evaluated the use of combined onset asynchrony and amplitude envelope cues. The results were compared to those with isolated onset asynchrony and envelope cues. Based on children's combined temporal cue use in sound segregation (Hall et al., 2008) and reported infant-adult cue-use differences in speech perception (Cabrera & Werner, 2017; Cooper & Aslin, 1994; Seidl & Cristia, 2008), it was hypothesized that the contribution of individual cues would be different for infants and adults and that infants may gain a greater benefit from combined cues than either cue in isolation.

## CHAPTER 2 INFANTS' USE OF ONSET ASYNCHRONY CUES

(Oster, M.-M. and Werner, L. A. (under review). Infants use onset asynchrony cues in auditory scene analysis. *Journal of the Acoustical Society of America*, Manuscript submitted for publication.)

### I. INTRODUCTION

A major developmental task in infancy and early childhood is the acquisition of language. Infants learn spoken language by listening to and interacting with their caregivers (e.g., Ramirez-Esparza, Garcia-Sierra, & Kuhl, 2017). However, typical daycare and home environments are noisy, requiring infants to segregate speech from competing sounds (Alston & James-Roberts, 2005; McAllister, Granqvist, Sjölander, & Sundberg, 2009). Prior research has shown that infants have greater difficulties than adults with this task, but the reasons are not well understood (for a review see Buss et al., 2012; Leibold, 2017).

Segregating competing sounds is a complex task. The sounds from competing sources arrive as a sound mixture at the ears. The combined sound is encoded in the periphery into a series of overlapping frequency bands. Frequency bands with similar spectral, temporal and spatial location characteristics are grouped; those with differing characteristics are separated. Auditory attention supports and guides this process. The acoustic characteristics that allow grouping and segregation are referred to as auditory cues (Bregman, 1990).

The ability to encode sound is mature early in life, and thus cannot account for infants' difficulties segregating competing sounds (for a review see Buss et al., 2012). However, whether infants are able to use auditory cues to group and separate frequency components is less clear. The results of several studies suggest that infants are able to use frequency separation (Demany, 1982; Winkler et al., 2003), harmonicity (Bendixen et al., 2015; Folland et al., 2015; Folland et al., 2012), and spatial location cues (Nozza, 1988). However, the results of these studies fail to provide a cohesive account of the development of sound segregation for several reasons.

First, only a limited number of auditory cues has been studied. Notably, temporal cues, some of the strongest cues in adults (Bregman, 1990), have not been systematically investigated. Furthermore, the auditory cue is rarely systematically varied in developmental studies. Instead, infants are presented with a cue that is large enough to allow adults to easily segregate the sounds involved (e.g., Bendixen et al., 2015; Winkler et al., 2003). Although this approach demonstrates that infants can use a specific cue, it does not provide a way to assess how efficiently infants use that auditory cue relative to adults. In fact, developmental effects in sound source segregation are typically not investigated directly. Most studies have not included infants at different ages, and the methods used preclude testing adults.

Another issue is that infants' ability to use auditory cues to segregate competing sounds is primarily investigated with non-speech stimuli (but see Leibold et al., 2013; Newman & Morini, 2017). The benefit of such stimuli is that they are easily controlled. However, for adults cue use with non-speech stimuli or synthetic speech does not always predict how auditory cues are used to segregate naturally produced speech (e.g., Nittrouer, 2004). This may be the case for infants as well.

Finally, electrophysiological methods are commonly used to investigate infants' ability to use an auditory cue to segregate sounds (e.g., Bendixen et al., 2015; Folland et al., 2015; Winkler et al., 2003). While such studies suggest that infants have access to a cue, they don't provide information about how infants use that cue.

The current study expands the number of isolated cues studied in infants by focusing on infants' use of one temporal cue, onset asynchrony. Onset asynchrony is a particularly strong cue for adults (Bregman, 1990), understandably because sounds from different sources rarely begin or end simultaneously. Its benefit is dependent on onset asynchrony duration (Hedrick & Madix,

2009; Lentz & Marsh, 2006; Valentine & Lentz, 2012). A small number of studies has investigated onset asynchrony cues in children. These studies indicate that introducing onset asynchrony improved tone detection in paradigms involving across-frequency processing, such as informational masking (Hall, Buss, & Grose, 2005; Leibold & Neff, 2007) and comodulation masking release (Hall et al., 2008), for both children and adults with little evidence of age-related improvement in the use of asynchrony cues. Although infants' use of onset asynchrony cues has not been systematically investigated, Bendixen et al. (2015) provided some evidence that newborns are sensitive to onset asynchrony. In this study, onset asynchrony plus a mistuning of one component of a harmonic complex lead to a different event-related potential morphology than mistuning alone. Although that is not evidence that the onset asynchrony contributed to the segregation of the mistuned harmonic, the results suggest that infants have access to this auditory cue.

Unlike most previous work, this study used speech sounds as stimuli. Although recent studies have used natural speech to investigate infants' ability to use talker gender to segregate competing speech, the stimuli used make it difficult to determine which cues were important for segregation (Leibold et al., 2013; Newman & Morini, 2017). For example, Newman & Morini (2017) played an ongoing speech stream and presented infants with short sentences from a talker of the opposite sex. In this study infants had access not only to voice differences but also to onset cues, offset cues and cues related to spectro-temporal variation. To avoid this problem, the current study used vowels. The spectro-temporal characteristics of vowels can easily be controlled and characterized, making it possible to investigate isolated auditory cues.

The current study addresses the remaining limitations discussed above by using a single behavioral method to test infants of different ages as well as adults. We developed an

experimental paradigm in which two different vowels are presented simultaneously and listeners are trained to respond to one specific vowel. While the paradigm here is similar to the double-vowel paradigm (e.g., Assmann & Summerfield, 1990), it differs from the double-vowel paradigm in several ways. First, naturally produced vowels from two different sex talkers are used to maximize the likelihood that infants will be successful in this task. Typically, double-vowel paradigms use synthetic vowels to control spectral cues to sound segregation. Second, concurrent vowels are presented at age specific target-to-masker ratios (TMR). Pilot work showed that infants require a positive TMR to prevent floor effects and adults require a negative TMR to prevent ceiling effects. Typically, double-vowel paradigms present the two concurrent vowels with a 0 dB TMR. Third, listeners are presented with a train of concurrent vowels and learn to respond when a specific target vowel is detected. Listeners in a traditional double-vowel paradigm are asked to identify the two vowels heard in a single presentation.

The aims of the current study were to investigate whether 3- and 7-month-old infants use onset asynchronies to segregate competing vowels and whether this ability improves with age as well as with onset asynchrony duration. To this end, two experiments were conducted. Experiment 1 investigated whether 7-month-old infants and adults were better able to segregate concurrent vowels with an onset asynchrony compared to vowels with synchronous onset. It was hypothesized that infants would receive a benefit from onset asynchrony, but that this benefit would be smaller than in adults. Onset asynchronies of 100 and 200 ms were chosen because results by Lentz and Marsh (2006) showed that adults tested in a similar task received a larger benefit from 200- than 100-ms asynchrony. It was predicted that increasing onset asynchrony duration would improve performance more for infants than for adults, given previous work showing that children require larger acoustic cues to achieve adultlike segregation (Sussman et

al., 2007). Experiment 2 was conducted to investigate whether 3-month-old infants use onset asynchrony cues to segregate concurrent vowels. The approach used in Experiment 1 was modified to assess the conditions under which 3-month-old infants could perform the task. It was predicted that more 3-month-old infants would be able to segregate concurrent vowels with an onset asynchrony than without and that more 3-month-olds would succeed with 200- than with 100-ms onset asynchrony.

## II. EXPERIMENT 1

### *A. Method*

#### *Design*

Participants heard a series of vowel pairs, temporally aligned, superimposed vowels spoken by a male and a female talker. They learned to respond when they heard a target vowel, /i:/ or /u:/, spoken by the male talker. A mixed design was employed. Each participant in the final sample provided data in both a baseline condition (synchronous vowels) and an asynchrony condition. Half of the subjects heard each target vowel. In the asynchrony condition, half of the participants heard the competing vowel begin 100 ms before the target vowel; and half heard the competing vowel begin 200 ms before the target vowel. The target-to-masker ratio (TMR) expected to produce a  $d' = 1$  in the synchronous onset condition for each age group was determined in pilot testing. All participants in an age group heard the same TMR. A benefit of asynchrony would be taken as an increase in  $d'$  in the asynchronous onset condition relative to the synchronous onset condition.

### *Participants*

A total of 44 7-month-old infants and 31 18- to 30-year-old adults participated. Inclusion criteria included full-term birth, passing newborn hearing screening (infants, per parental report) or passing hearing screening on the test day (adults), no family history of congenital hearing loss, no recent treatment for otitis media, no risk factors for hearing loss, passing tympanometric screening with a 1-kHz probe tone (infants) or 226 Hz probe tone (adults) on each test day, and no prior experience in psychoacoustic experiments. Because the stimuli were American English vowels, only native English speakers or infants from homes where English is the dominant language were included. Participants were recruited through the University of Washington Communication Studies Participant Pool, a facility maintaining a database of individuals willing to participate, or to have their child participate, in research. Some adults were recruited through flyers and newspaper ads. All participants were paid for their participation. Infants completed testing in 1 to 3 visits, adults completed testing in 1 visit.

Data were only included in the analyses if a participant completed both experimental conditions with  $d' \geq 0.8$ <sup>1</sup> in at least one condition, resulting in a final sample of 24 infants (mean age 30 wk, SD = 2.3) and 24 adults (mean age 22 yr, SD = 3.7). The data of 27 participants were excluded from analyses. Of those, 15 infants and 8 adults reached training criteria but did not complete two conditions with at least one  $d' \geq 0.8$ . Another 5 infants and 3 adults did not reach training criteria.

### *Stimuli*

The American-English vowels /i:/, /u:/, /ʊ/, /æ/, /ɛ/, /ɔ:/ and /ɪ/ were the stimuli. They were recorded by one male and one female native speaker in a professional recording studio

---

<sup>1</sup> If an observer responded “yes” randomly on each trial, the probability of obtaining a  $d' \geq 0.8$  on 30 trials is less than 0.05, assuming a binomial process with  $p = 0.5$ .

using a high-quality condenser microphone. Recordings were digitized at a resolution of 32 bits and a sampling rate of 44.1 kHz. The recordings were processed using PRAAT (Boersma & Weenink, 2017) to produce the final stimuli. The duration of each vowel was shortened to 400 ms. The pitch contour from one representative vowel from each talker was extracted using the PSOLA algorithm and applied to all of the vowels of that talker. The fundamental frequency (F0) of vowels ranged from 83 to 123 Hz for the male talker and from 219 to 235 Hz for the female talker. Figure 1 shows the position of each vowel in F1/F2 space. The male /i:/and /u:/ were chosen as targets because, as Figure 1 indicates, the 1<sup>st</sup> and 2<sup>nd</sup> formants of these vowels are separated from those of the masker vowels. This is expected to reduce, but not eliminate, energetic masking of the target formants by masker formants. The root-mean-square (RMS) amplitudes of each talker's vowels were equated using the scale intensity function in PRAAT. 15-ms on- and off-ramps were included. The level of the masker (female) vowels was set at 60 dB SPL for all listeners in all conditions. The level of the target (male) vowels was set to 75, 72, 60, 50 or 40 dB SPL, depending on the test phase and age of the listener.

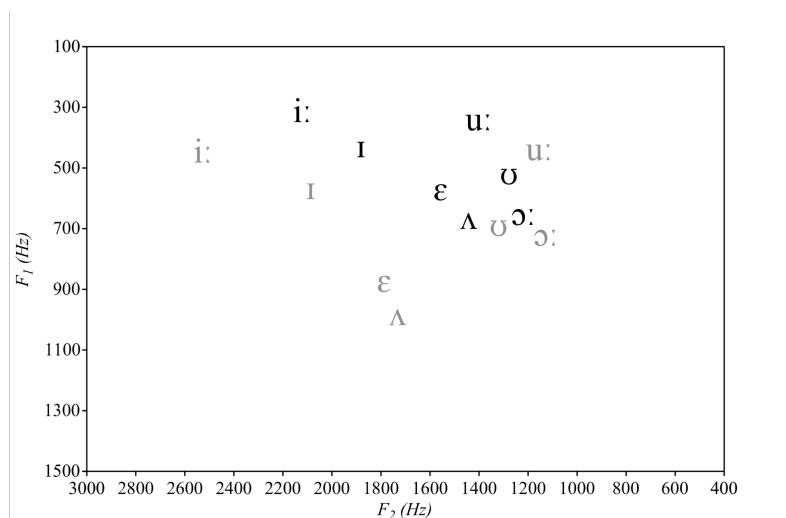


Figure 2-1: F1/F2 space for the vowels used in experiments 1 and 2.

F1 and F2 were obtained at 100 ms for each vowel. Male talker vowels are plotted in black, female talker vowels in grey.

The male and female vowels were combined offline to create the final stimulus set. The two vowels in a pair were always different, and the masker vowel was never the target vowel, e.g. when /u:/ was the target vowel, it was not included as a masker in any vowel pair. Therefore, of the 49 possible vowel pairs, 7 were eliminated because they contained matching vowels and 6 were eliminated because they contained the target vowel in the masker. This left a total of 36 vowel pairs: 30 with no target and 6 with a target vowel. In the synchronous onset condition, the vowels were aligned with simultaneous onset and offset. In the asynchronous onset conditions, the target-talker vowels were shortened to the appropriate durations by deleting either 100 or 200 ms at the end of the vowel. Off ramps of 15 ms were imposed on the shortened vowel. The levels of these shortened target-talker vowels were scaled to 75, 72, or 40 dB SPL, depending on the test phase and age of the listener. Zeroes were added to the beginning of the file to equate the length of the target-talker and masker vowels before the files were combined, resulting in simultaneous offsets. This process was completed for each TMR used in the experiment.

### *Procedure*

The observer-based psychophysical procedure was used to assess infant sensitivity (Werner, 1995). Infants were seated on their caregiver's lap in a single-walled sound-attenuated booth. An assistant was seated to the infant's left and manipulated quiet toys to keep the infant engaged during the test session. A tower with dark plexiglass boxes containing mechanical toys and a video monitor, the "visual reinforcer", was located to the infant's right. The observer was seated outside of the sound booth and watched the infant through an observation window. Throughout the test session, the double-vowel pairs were presented repeatedly with an ISI of 1200 ms from a loudspeaker (JBL; Professional Control 1) located at the front of the sound booth. The caregiver, assistant and observer heard a continuous stream of double-vowels

presented over insert earphones or headphones to prevent them from hearing the sounds presented to the infant. When the infant appeared attentive, the observer initiated a trial. On each trial, the infant heard the target vowel (“target trial”) or a non-target vowel (“no-target trial”), paired with a masker vowel. Target and no-target trials occurred in pseudorandom order, with constraints as described below. Target and non-target vowel pairs were randomly selected from the available pairs of each type. The observer was blind to trial type and judged whether or not the target vowel had been presented on each trial, based on the infant’s behavior. The observer had 4 s after trial onset to record a response. The observer received feedback after each trial. When the observer correctly identified a target trial, the mechanical toy or video was turned on for 4 s, to reinforce the infant’s response. The assistant in the booth also delivered “social reinforcement” to the infant when the toy or video was activated. When the observer’s judgment was incorrect or when the observer correctly identified a no-target trial, no reinforcement was provided.

Each session began with a demonstration phase to familiarize the infant with the target vowel and the procedure. Vowel pairs were presented with simultaneous onsets. The TMR was 0 dB for all non-target vowel pairs; the TMR was +15dB for target vowel pairs. The probability of a target trial was 0.8. The visual reinforcer was activated 2 s after a target vowel pair was presented, regardless of the observer’s response. The demonstration phase continued until the infant-observer team was correct on 4 of the last 5 target trials and 1 no-target trial, until the infant’s state precluded further testing, or until 20 trials had been presented.

The session continued with a training phase to teach the infant that a response was required to activate the visual reinforcer and to teach the observer what infant responses to target vowels might be. Vowel pairs were still presented with simultaneous onsets, but the TMR was

+15 dB for all vowel pairs. The probability of a target trial was 0.5. The reinforcer was activated only when the observer correctly identified a target trial. The training phase continued until the infant-observer team correctly identified 4 of the last 5 target trials and 4 of the last 5 no-target trials (training criterion), until the infant's state precluded further testing, or until 40 trials had been presented.

Infants who reached training criterion in the training phase completed two testing phases, one in the synchronous onset condition and one in an asynchronous onset condition, order counterbalanced across participants. During the testing phases the TMR was +12 dB for all vowel pairs on target and no-target trials. Thirty-six trials were presented in 2 blocks. One block contained 7 target trials and 7 no-target trials; the other contained 8 target and 8 no-target trials<sup>2</sup>. In addition, both blocks contained 3 probe trials, target trials with +15 dB TMR. Probe trials were included to ensure that infants were still engaged in the task. Within each block, the trial types were presented in random order. The block order was counterbalanced across participants. The infant only received reinforcement when the observer correctly identified a target or a probe trial. Testing was discontinued if the observer missed 3 probe trials or if the infant's state precluded further testing. Otherwise testing continued until all 36 trials were completed.

If an infant did not reach criterion in the training phase during a visit, the next visit began with a demonstration phase. If an infant completed the training phase and/or one testing phase, subsequent visits began with a few "reminder" trials from the demonstration and training phases before beginning the next testing phase. Some infants completed training and one testing phase within one visit and the other testing phase in a separate visit. However, most infants completed training and the two testing phases in separate visits.

---

<sup>2</sup> In the block with 7 target vowel pairs, the target vowel was combined with the competing vowels /æ, ʊ, ɔ:, ɪ, æ, ʊ, and i: or u: (whichever vowel was not the target vowel)/. In the block with 8 target vowel pairs, the target vowel was combined with the competing vowels /ε, ɪ, æ, ʊ, ɔ:, ε, ɪ, and i: or u:(whichever vowel was not the target vowel)/.

Adults were tested in a similar procedure. Adults were seated alone in the test booth. They were instructed to raise their hand when they heard the sound that “makes the toy come on”. Adults completed all phases in a 1-hour visit with a short break between the two testing phases. The stimuli in the demonstration phase were the same for infants and adults. However, for adults the TMR was -10 dB in the training phase and -20 dB TMR in the testing phases.

### ***B. Results***

The measure of performance was  $d'$ . Perfect hit and false-alarm rates were adjusted by  $1/n$ , where  $n$  is the number of trials. Target vowel, order of condition, and order of experimental blocks did not affect  $d'$ . Thus, results were collapsed over target vowel and order of presentation.

Figure 2 plots  $d'$  in the synchronous and asynchronous onset condition for individual subjects in each age group. The overall level of performance is variable, which is not surprising given that listeners in each age group were tested with a fixed TMR. While the benefit of asynchrony is also variable, 14 of 24 infants and 15 of 24 adults performed better in the asynchronous than in the synchronous onset condition.

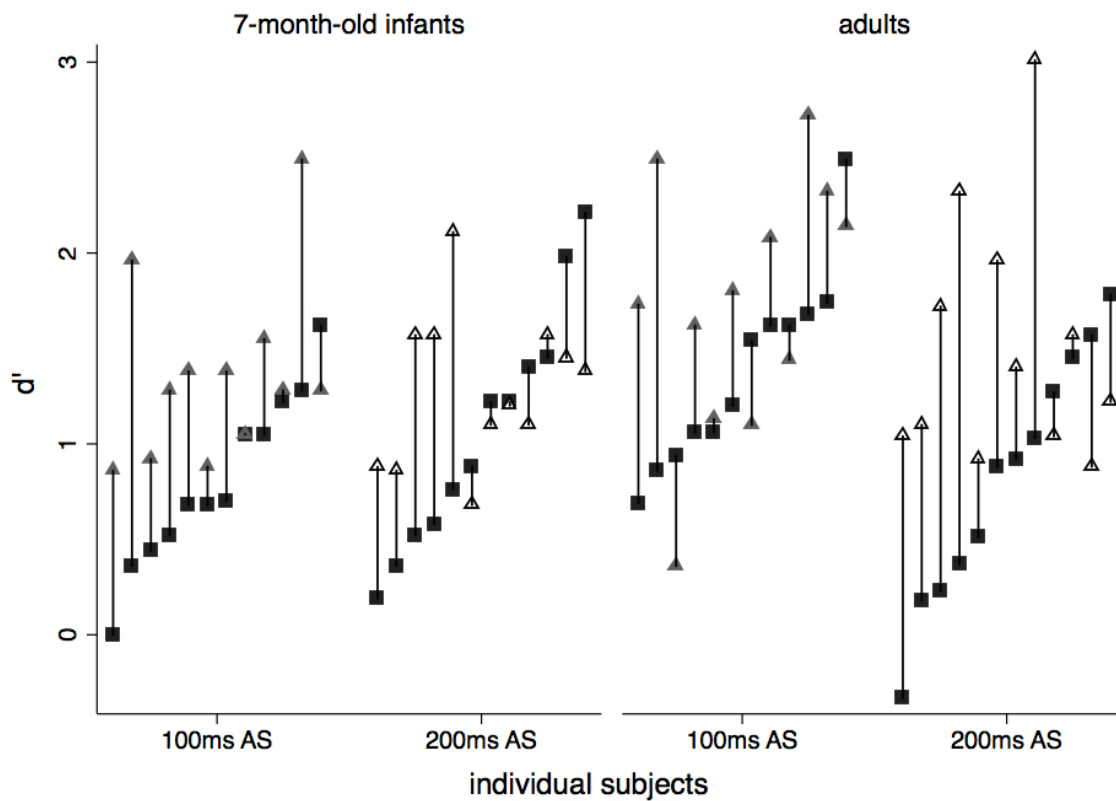


Figure 2-2 Individual  $d'$  scores for infants and adults.

Scores are ordered by performance in the synchronous onset condition and divided by onset-asynchrony duration. Squares represent  $d'$  in the synchronous onset condition, triangles represent  $d'$  in the asynchronous onset condition. The lines connect data points from the same participant.

Figure 3 shows average  $d'$  for infants and adults as a function of condition. As intended, average  $d'$  in the synchronous onset condition was approximately 1, with a range of 0.8 to 1.4, across groups. Comparing performance between the synchronous onset and onset asynchrony conditions, both infants and adults performed better with asynchronous than with synchronous onsets for both asynchrony durations. No differences in performance are apparent between the 100- and 200-ms onset asynchrony conditions within age group.

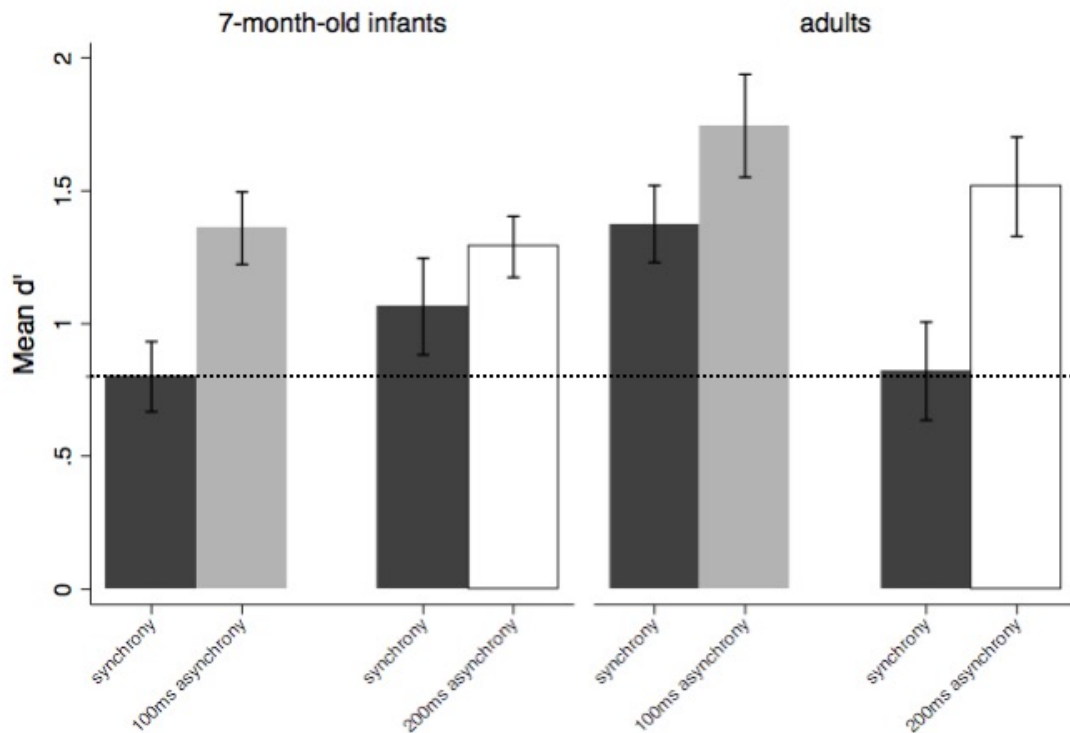


Figure 2-3 Average  $d'$  in each condition by age group.

The dotted line represents above chance performance. Error bars represent +/- 1 standard error of the mean

To statistically evaluate these observations a mixed models analysis was performed using Generalized Estimating Equations (GEE) assuming an exchangeable correlation structure and robust standard errors. The full model included the dependent variable  $d'$  and the independent variables asynchrony (synchronous onset, asynchronous onset), duration (100 ms, 200 ms), age (7-month-old infant, adult), and all possible interactions. The results are summarized in Table 1. All three main effects reached statistical significance, but because the interaction between age and onset asynchrony duration also reached statistical significance, this interaction is interpreted

first. Post-hoc analysis showed that for infants, the effect of duration was nonsignificant ( $z = 0.63$ ,  $p = 0.529$ ) and that for adults, the effect of duration just missed statistical significance ( $z = -1.91$ ,  $p = 0.056$ ). Note that adult performance was actually better at the 100- than at the 200-ms onset asynchrony. Thus, although the main effect of duration was significant, it is clear from the preceding analysis that duration only influences performance in adults in an unexpected direction. In contrast, the main effect of age seems to hold in all conditions. Given that the age effect was not specific to the asynchrony conditions, there is no evidence that the benefit of onset asynchrony is different between infants and adults.

Table 2-1 Results of mixed models GEE statistical analysis of  $d'$ .

Predictor	$B$	SE $B$ (robust)	$z$	$p$	95% Confidence interval
age	-0.53	0.19	-2.85	0.004	-0.9 – -0.17
asynchrony	0.41	0.19	2.14	0.032	0.03 - 0.78
duration	-0.51	0.24	-2.12	0.034	-0.98 - -0.04
age x asynchrony	0.15	0.23	0.63	0.529	-0.32 - 0.63
age x duration	0.77	0.32	2.38	0.017	0.14 - 1.4
age x asynchrony x duration	-0.63	0.41	-1.53	0.126	-1.43 - 0.18

### *C. Discussion*

The results of Experiment 1 show that 7-month-old infants gain a similar benefit from onset asynchrony as adults in the segregation of concurrent vowels. In fact, infants and adults were remarkably similar in their use of onset asynchrony. The same number of infants received a benefit from the asynchrony cue as adults. The variability in performance is similar between the age groups. Within each age group, increasing duration does not improve performance. We believe that the trend for decreased performance between 100- and 200-ms onset asynchronies for adults is likely the result of sampling error.

Not all subjects showed a benefit from onset asynchrony. It should be noted that variability in benefit is commonly noted in the literature (Hall et al., 2008; Hill & Darwin, 1996; Leibold & Neff, 2007; Lentz & Marsh, 2006; Singh & Bregman, 1997). It might also be the case that the TMR used was not sufficient to support detection in either condition for some listeners. In addition, day-to-day (or minute-to-minute) variability in performance is a natural feature of untrained listeners, and the small number of trials used here exacerbates the problem. Thus, the failure to find benefit of onset asynchrony in some listeners probably doesn't mean that those listeners are unable to use onset asynchrony cues.

Contrary to expectation, asynchrony duration did not lead to an improvement in performance in either age group. There are differences in stimuli and procedures between this study and those demonstrating a duration effect. However, it should be noted that while performance in concurrent vowel segregation generally improves with improving onset asynchrony (Lentz & Marsh, 2006), not all studies find an improvement between these two asynchrony durations (Valentine & Lentz, 2012). It is possible that the lack of a positive duration effect in this study is due to the high variability in performance, small sample size, and small number of trials.

### III. EXPERIMENT 2

While Experiment 1 focused on 7-month olds and adults, Experiment 2 evaluated onset asynchrony cue use in 3-month-old infants. Given the dramatic changes in the auditory system in the first 6 months of life (Eggermont & Moore, 2012), and the fact that newborns have access to but may not use onset asynchrony cues (Bendixen et al., 2015), it might be expected that infants younger than 7 months of age do not use onset asynchronies to the same degree as older infants

or adults. Furthermore, it might be predicted that infants younger than 7 months require longer asynchronies to segregate concurrent vowels.

Unfortunately, initial test results showed that 3-month-old infants were unable to complete the test phases that provided estimates of  $d'$  in each condition. Of the 22 3-month-old infants initially tested, only 9 were able to complete both the synchronous and asynchronous onset cue conditions with above-chance performance in at least one condition. However, 14 reached criterion in the training phase in the synchronous onset condition. Thus, to get information about 3-month-olds' use of onset asynchrony cues in vowel segregation, different groups of infants were tested with synchronous vowel onset, 100-ms onset asynchrony and 200-ms onset asynchrony, respectively. The proportion of infants reaching the training criterion in each condition was taken as a measure of task difficulty. If more infants are able to reach criterion in an asynchronous onset condition compared to the synchronous onset condition, this was taken as evidence that onset asynchrony provides a benefit in vowel segregation. This approach is not unlike that used in many other developmental studies (e.g., Cabrera & Werner, 2017; Kuhl, 1979; Polka & Bohn, 2003). It was hypothesized that a larger proportion of infants would be able to reach criterion with asynchronous vowel pairs. It was furthermore hypothesized that a larger proportion of infants would reach criterion with a 200-ms onset asynchrony compared to a 100-ms onset asynchrony.

## ***A. Methods***

### *Participants*

A total of 67 3-month-old infants (mean age: 12 weeks,  $SD = 0.7$ ) participated in this experiment. The 22 infants who were initially trained in the synchronous onset condition were included. In addition, 22 infants provided data in the 100-ms asynchronous onset condition and

23 infants provided data in the 200-ms asynchronous onset condition. The inclusion criteria were identical to those of Experiment 1. An additional two infants were excluded, 1 infant due to excessive fussiness on both test days (synchronous onset condition) and 1 infant who failed tympanometry screening (100-ms asynchronous onset condition).

### *Stimuli*

The stimuli used in Experiment 2 were the same as those used in the demonstration and training phases for infants in Experiment 1, except that vowel pairs in the training phase of Experiment 2 could have synchronous onset, a 100-ms onset asynchrony, or a 200-ms onset asynchrony.

### *Procedure*

The demonstration and training phase procedures described in Experiment 1 were used in Experiment 2, with the exception that infants were assigned to one of three different asynchrony conditions in the training phase: synchronous onset, 100-ms onset asynchrony, or 200-ms onset asynchrony.

A test session continued until the infant reached criterion; the infant became unresponsive due to fussiness, hunger, or sleepiness; or until the maximum number of trials in a phase had been presented. If a session was aborted prior to reaching criterion or the maximum number of trials, a second session was attempted in the same visit, if the infant was willing. This procedure was followed on each of 2 visits to the lab. An infant was only judged to have failed to reach criterion if at least 2 sessions were attempted and only if the 2 sessions were attempted at different visits.

## ***B. Results***

Fourteen of 22 infants (64%) reached criterion in the synchronous onset condition, while 21 of 22 infants (95%) reached criterion in the 100-ms asynchronous onset condition and 20 of 23 infants (87%) in the 200-ms asynchronous onset condition.

Three analyses were performed to evaluate 3-month-old infants' use of onset asynchrony cues. The first analysis tested whether the number of infants who successfully reached criterion in each condition was greater than would be expected by chance. The second analysis tested whether infants in the asynchronous onset conditions were more likely to reach criterion than infants in the synchronous onset condition. The third analysis tested whether the proportion of infants reaching criterion differed between the two asynchronous onset conditions.

In the first analysis, a simulation of 10,000 infants was performed to determine the proportion of infants who would reach criterion if the observer guessed on every trial. The simulation followed the same rules as the test procedure in terms of target probability, maximum number of trials, and maximum number of sessions attempted. Two parameters were determined from the test data and included in the simulation: 1) the rate at which the observer judged that a target trial had occurred in each phase, regardless of trial type and how many trials were completed in a session, and 2) the probability that a session was aborted due to infant state. The "target" response rate was 0.67 in the demonstration phase and 0.59 in the training phase. The proportion of aborted sessions was 0.12 for 1<sup>st</sup> sessions, 0.09 for 2<sup>nd</sup> sessions, 0.04 for 3<sup>rd</sup> sessions, and 0 for 4th sessions. Based on this simulation, the proportion of infants expected to reach criterion by chance was 0.44. One-sided exact binomial tests showed that the proportion of infants reaching criterion was greater than expected by chance in the 100- and 200-ms asynchronous onset conditions ( $p < 0.001$ ). In the synchronous onset condition, the one-sided

exact binomial test just missed statistical significance ( $p = 0.051$ ), suggesting that segregating synchronous vowels was difficult for 3-month olds.

The second analysis evaluated whether onset asynchrony provided a benefit in concurrent vowel segregation. A one-tailed statistical test of proportions showed that the proportions of infants reaching criterion in the 100- and 200-ms asynchrony condition were significantly greater than in the synchronous onset condition (100 ms:  $z = -3.06$ ,  $p = 0.0054$ ; 200 ms:  $z = -1.82$ ,  $p = 0.0344$ ).

The third analysis evaluated whether asynchrony duration influenced concurrent vowel segregation in 3-month-old infants. A one-tailed statistical test of proportions showed that the proportion of infants reaching criterion in the 200-ms asynchronous onset condition was not greater than in the 100-ms asynchronous onset condition ( $z = 1.71$ ,  $p = 0.9568$ ).

### ***C. Discussion***

The results of Experiment 2 suggest that 3-month old infants benefit from onset asynchrony cues in concurrent vowel segregation. Contrary to expectation, the duration of onset asynchrony did not influence the proportion of 3-month olds reaching criterion. The potential reasons for the failure to find a duration effect are the same as those discussed in Experiment 1. It is also possible that the proportion of infants reaching criterion is not a sensitive enough measure to detect differences between 100- and 200-ms onset asynchrony durations.

A direct comparison between 3-month olds, 7-month olds and adults is not possible, because the measure of sensitivity used in Experiment 1,  $d'$ , is different from that in Experiment 2, the proportion of infants reaching criterion. However,  $d'$  in the synchronous onset and asynchronous onset conditions are available for nine 3-month-old infants. This subset of 3-month olds who were able to provide data in both experimental conditions is likely biased toward the

best performers, given that segregation of vowels with synchronous onset appears to be a difficult task for 3-month-old infants. Nonetheless, for this subset of infants, average  $d'$  in the asynchronous onset conditions (mean = 1.13, SD = 0.27) was higher than in the baseline condition (mean = 0.88, SD = 0.34), suggesting a benefit of the onset asynchrony cue. The benefit 3-month olds experienced appears slightly smaller than in 7-month olds and adults, but for obvious reasons it is difficult to draw an strong conclusions about age-related improvement on the basis of this comparison. Clearly, whether sensitivity to onset asynchrony develops between 3 and 7 months of age cannot be determined without further procedural adjustments to allow a greater number of 3-month-old infants to provide sensitivity data.

#### IV. GENERAL DISCUSSION

The current study provides the first behavioral evidence that 3- and 7-month-old infants use onset asynchrony cues to group and segregate the components of speech sounds. In fact by 7 months of age infants use onset asynchrony similarly to adults when performance is equated in synchronous vowel segregation. Although not enough data were obtained from 3-month-old infants to directly compare age groups, the available sensitivity scores suggest that 3-month-old infants may be less efficient than both 7-month olds and adults at using onset asynchrony cues. Another outcome of this study was that increased duration did not lead to an improvement in performance within any age group. It seems likely that this result is due to the presence of additional cues to sound source segregation, those based on voice characteristics. However high variability in performance, small sample size, and small number of trials probably also contribute to the lack of duration effect. Similarly, the nonsignificant trend of decreased performance with 100- and 200-ms onset asynchronies in adults is most likely a result of sampling error.

By combining the results of this study with those of prior studies, we might speculate on the possible developmental timeline of onset asynchrony use in sound segregation. Newborns have access to onset asynchrony, but may not yet use this cue to segregate sounds (Bendixen et al., 2015). By 3 months of age infants use onset asynchrony to segregate concurrent vowels. However, the use of onset asynchrony may become more efficient between 3 and 7 months. By 7 months of age, infants seem to gain as much benefit from onset asynchrony cues as adults. The large variability observed in this and other studies may indicate that the use of onset asynchrony cues may develop more slowly in some children (Hall et al., 2005; Hall et al., 2008; Leibold & Neff, 2007). Thus, while onset asynchrony does not appear to be a major contributor to infants' and children's difficulties segregating competing sounds, it may contribute for some infants and children.

While the current study demonstrated that 3- and 7-month-old infants can segregate concurrent vowels, it should be noted that infants required a +12 dB TMR to achieve a  $d'$  of 1 in the synchronous onset condition, whereas adults only required a -20 dB TMR. The fact that infants required a greater TMR is not surprising. For example, infants consistently require a greater TMR than adults in tone or speech-sound detection in broadband noise (Nozza et al., 1990; Nozza & Wilson, 1984). The age-related difference in required TMR increases with increasing variability of the stimuli. However, even in informational masking paradigms, where masking cannot be explained by peripheral processes alone and large developmental effects are observed, the age-related TMR difference is not as great as that observed in the current study (Leibold & Werner, 2006). However, the age-related TMR difference in the current study is similar to that reported in other studies of speech-on-speech detection. For example, Leibold et

al. (2013) found that the average threshold TMR difference between infants and adults was 27.5 dB SPL, compared to 32 dB SPL here.

The large age-related threshold TMR difference indicates that infants and adults use different cues to segregate speech sounds in the absence of temporal cues. Two recent studies have suggested that infants do not use voice characteristic cues to segregate speech from speech. Leibold et al. (2013) found that while adults' detection thresholds improved when the target and competing speech were produced by different-sex talkers, infants' detection thresholds were not influenced by the sex mismatch. Similarly, Newman and Morini (2017) found that 16-month olds' word recognition in a speech masker was not influenced by a talker-sex mismatch between the target and masker. That infants use temporal cues, but not voice characteristic cues, is consistent with the idea that infants listen selectively in time (Werner, Parrish, & Holmer, 2009), but not in frequency (Bargones & Werner, 1994). If infants do not use voice characteristic cues, then infants in the synchronous onset condition would have to rely on intensity cues to segregate competing vowels. Since their intensity resolution is immature (Bull, Eilers, & Oller, 1984; Sinnott & Aslin, 1985), this might provide yet another explanation for the large positive TMR required by infants. However, until the cues to sound segregation have been investigated in isolation—for example, by using synthetic vowels—it is not possible to determine which cues infants use.

In summary, these results show that by 3 months of age, infants can use onset asynchrony in the segregation of competing vowels, but hint that younger infants may be less efficient at using this cue than adults. Nonetheless, to the extent that immature sound source segregation contributes to infants' difficulties processing speech in the presence of competing sounds, an inability to use onset asynchrony is probably not a major contributor.

## CHAPTER 3 INFANTS' USE OF ENVELOPE CUES

### I. INTRODUCTION

Speech processing in the presence of competing sounds is immature in infants and develops well into childhood (Leibold, 2017). The ability to segregate sound sources depends in part on accurate auditory encoding. Since auditory encoding is largely mature by 6 months of age (Buss et al., 2012), it cannot explain infants' and children's difficulties dealing with competing sounds. However, sound source segregation also depends on specialized bottom-up and top-down mechanisms. Bottom-up mechanisms capitalize on the fact that frequency components emitted from the same sound source are related to each other in their spectral, temporal and location characteristics. For example, all frequency components from one sound source start and stop together and carry similar changes in amplitude over time. The auditory system uses these auditory cues to group frequency components that have similar acoustic characteristics and to segregate those that differ (Bregman, 1990). This process is supported and guided by top-down mechanisms, such as auditory attention (Shinn-Cunningham, 2008). Hence, the prolonged development of the ability to separate competing sounds may be related to immaturities in the mechanisms underlying sound source segregation. To evaluate this possibility the current study focuses on infants' ability to use auditory cues in sound source segregation.

Infants' use of auditory cues in sound source segregation has received limited attention to date. The existing studies have shown that infants can segregate competing sounds that differ in frequency separation and harmonicity (Bendixen et al., 2015; Demany, 1982; Folland et al., 2015; Folland et al., 2012; Winkler et al., 2003), as well as location cues (Nozza et al., 1990). Oster & Werner (under review) recently showed that infants can use a temporal cue, onset

asynchrony, to segregate sounds. However, variations in the temporal envelope, another temporal cue, has not been studied in the context of sound segregation in infants.

It is clear from the adult literature that amplitude modulation improves detection, discrimination, and speech reception in the presence of competing sounds. The majority of these studies have introduced modulation in the masker (Bernstein & Grant, 2009; Brungart, 2001; Grimault, Bacon, & Micheyl, 2002; Nie & Nelson, 2015). Although improvements in performance with masker modulation might be due to envelope cue-based sound segregation (e.g., Hall et al., 2006; Hall et al., 2008), they can also be explained by listeners' glimpsing the target during masker minima or by other mechanisms (Bernstein & Grant, 2009; Borrill & Moore, 2002; Brungart, 2001; Cooke, 2006; Fogerty, Xu, & Gibbs, 2016; Füllgrabe, Berthommier, & Lorenzi, 2006; Miller & Licklider, 1950). Fewer studies have examined the effect of target modulation, but those studies demonstrate that adults can use differences between amplitude envelopes to segregate competing sounds (Moore & Gockel, 2012; but see Summerfield et al., 1992). However, how effective differences in modulation are appears to depend on a complex interaction of envelope cues and spectral characteristics of competing sounds (Hall et al., 2006; Hall et al., 2008).

While infants' use of envelope cues in sound source segregation has not been studied, results from related studies suggest that infants and children may be limited in their use of such cues. First, infants seem insensitive to amplitude modulation (AM), requiring greater modulation depth to detect AM, despite adultlike temporal resolution (Horn et al., 2016; Werner, 2006; Werner et al., 2001). Second, infants' and young children's detection and discrimination improves less as a result of masker modulation than adults' (Grose & et al., 1993; Hall, Buss, & Grose, 2014; Hall, Buss, Grose, & Roush, 2012; Hall, Grose, & Dev, 1997; Stuart, 2005; Stuart, 2008; Werner,

2013). That envelope cues may not be salient cues for infants is further supported by the finding that infants do not notice asynchrony between the auditory and visual envelopes of speech (Lewkowicz & García Coll, 2010), and that differences in envelopes only lead to small improvements in performance in 4- to 11-year-old children (Hall et al., 2008). Note, however, that Cabrera and colleagues (Cabrera, Bertoncini, & Lorenzi, 2013; Cabrera & Werner, 2017) showed that infants use envelope information as well as adults to discriminate between consonants in quiet and in noise.

The ability to use envelope cues in sound segregation may be particularly relevant for infants and children with cochlear implants (CIs). While spectral and location cues are degraded in CIs, temporal cues, including envelope cues, are well represented. The utility of envelope cues in sound segregation in adult CI listeners has been demonstrated (Chatterjee, Sarampalis, & Oba, 2006; Pierzycki & Seeber, 2014). If the ability to use envelope cues in sound source segregation is late to develop, special efforts may be necessary to allow infants and children with CI to segregate sounds. Thus, understanding the development of envelope cue use is of some importance.

The current study evaluates the development of envelope cue use in infants. To this end, two experiments were conducted. The first experiment investigated envelope cue use in 7-month-old infants and adults. It was hypothesized that 7-month-olds would segregate competing sounds with different amplitude envelopes better than competing sounds with similar envelopes, but that this effect would be smaller than in adults. The second experiment investigated envelope cue use in 3-month-old infants. Since the first 6 months of life are a time of rapid auditory development (Eggermont & Moore, 2012), it was of interest to investigate whether differences would be observed in the use of envelope cues in 3- and 7-month-old infants.

## II. EXPERIMENT 1

### *A. Method*

#### *Design*

Experiment 1 employed a mixed design to measure the effect of envelope cues. Listeners heard a train of superimposed vowel pairs. One vowel in each pair was spoken by a man, the other was spoken by a woman. Listeners learned to respond to one of two target vowels: male /i:/ or male /u:/. An equal number of listeners heard each target vowel. Participants in the final sample provided data in two experimental conditions: the baseline condition, in which the amplitude envelopes of the vowels in a pair were similar, and the envelope cue condition, in which the amplitude envelopes of the vowels in a pair were different. The vowel pairs were presented at age specific target-to-masker ratios (TMR) that produced a  $d' = 1$  in the baseline condition in pilot testing. An increase in  $d'$  in the envelope cue condition relative to the baseline condition was interpreted as a benefit of the envelope cues.

#### *Participants*

A total of 12 7-month-old infants (mean age: 30 weeks, SD = 1.8) and 13 young adults (mean age: 21 years, SD = 2.7) provided data in both experimental conditions. Inclusion criteria included normal hearing (infants: passing newborn hearing screening per parent report, adults: passing hearing screening at 20 dB HL at octave frequencies between 250 and 4000 Hz), normal development, no family history of congenital hearing loss, no risk factors for hearing loss, no recent treatment for otitis media, passing tympanometry screening on each test date (1-kHz probe tone for infants, 226 Hz probe tone for adults), English as the dominant language in the home, and no prior experience in psychoacoustic experiments. Participants were recruited through the Communication Studies Participant Pool at the University of Washington, which maintains a

database of individuals who are willing to participate in research or who are willing to have their child participate in research. Some adult participants were recruited through flyers and newspaper announcements.

An additional 13 participants were enrolled in this study but excluded for the following reasons: failed tympanometry screening (2 infants), inability to learn the task (4 infants), inability to complete 2 experimental conditions (2 infants), below chance performance in both experimental conditions<sup>3</sup> (2 infants and 2 adults), and ceiling performance in the baseline condition (1 adult).

### *Stimuli*

Sound source segregation in infants has been primarily investigated using non-speech stimuli (e.g., Bendixen et al., 2015; Demany, 1982; Folland et al., 2015; Folland et al., 2012; Winkler et al., 2003). Although those stimuli allow for precise characterization and manipulation of auditory cues, they do not always accurately predict how listeners use auditory cues to segregate speech from competing sounds (Nittrouer, Lowenstein, & Tarr, 2013; Sussman et al., 2007). For that reason, speech sounds were used as stimuli in this study, but the speech sounds used were relatively manageable vowels.

The 7 American-English vowels /i: u: ɔ æ ε ɔ: ɪ/ were used to generate double-vowel pairs. The vowels were recorded by a male and by a female talker in a professional recording studio using a high-quality condenser microphone. Recordings were digitized at 32 bits and a sampling rate of 44100 Hz. Only vowel tokens with relatively flat envelopes were selected for further processing. Using PRAAT (Boersma & Weenink, 2017), all vowels were shortened to 400 ms. The pitch contour of one vowel from each talker was extracted using the PSOLA

---

<sup>3</sup> If an observer responded “yes” randomly on each trial, the probability of obtaining a  $d' \geq 0.8$  on 30 trials is less than 0.05, assuming a binomial process with  $p = 0.5$ .

algorithm and applied to all vowels from that talker, respectively. The range of fundamental frequencies (F0) was 219 to 235 Hz for the female talker and 83 to 123 Hz for the male talker.

The first two formant frequencies of the vowels are listed in Table 1.

The vowels from each talker were scaled in root-mean-square (RMS) amplitude with 15-ms on- and off-ramps using the scale intensity function in PRAAT. The level of the female vowels was always 60 dB SPL. The level of the male vowels was 40, 50, 60, 72, or 75 dB to generate TMRs of -20 and -10 dB for adults and +12 and +15 dB for infants.

Table 3-11<sup>st</sup> and 2<sup>nd</sup> formant (F1, F2) of all vowels from each talker.

		/i:/	/u:/	/ɔ:/	/ɛ/	/ɪ/	/æ/	/ʊ/
Female talker	F1	449 Hz	440 Hz	724 Hz	874 Hz	564 Hz	983 Hz	689 Hz
	F2	2518 Hz	1129 Hz	1115 Hz	1782 Hz	2068 Hz	1733 Hz	1308 Hz
Male talker	F1	316 Hz	343 Hz	653 Hz	571 Hz	430 Hz	666 Hz	517 Hz
	F2	2113 Hz	1390 Hz	1189 Hz	1553 Hz	1854 Hz	1423 Hz	1274 Hz

Using these stimuli, vowel pairs were created by superimposing 2 different vowels, one from each talker. The male /i:/ and male /u:/ were selected as target vowels because their first two formants are separated from those of the masker (female) vowels, reducing energetic masking (see Table 1). The target vowel was never included among the competing vowels; that is, if /i:/ was the target vowel, /i:/ was not included in the masker vowels. In this way 36 vowel pairs were created for the baseline condition at each TMR, 30 vowel pairs without a target vowel and 6 vowel pairs with the target vowel.

For the envelope cue condition, the female vowels were the same as in the baseline condition. However, the amplitude envelopes of the male talker's vowels were modified to represent those found in infant-directed speech. To this end, the male talker was instructed to record vowels as if speaking to an infant. From these tokens, vowels with 3 different amplitude

envelopes were selected and shortened to 400 ms. The 3 amplitude envelopes are schematically represented in Figure 1.

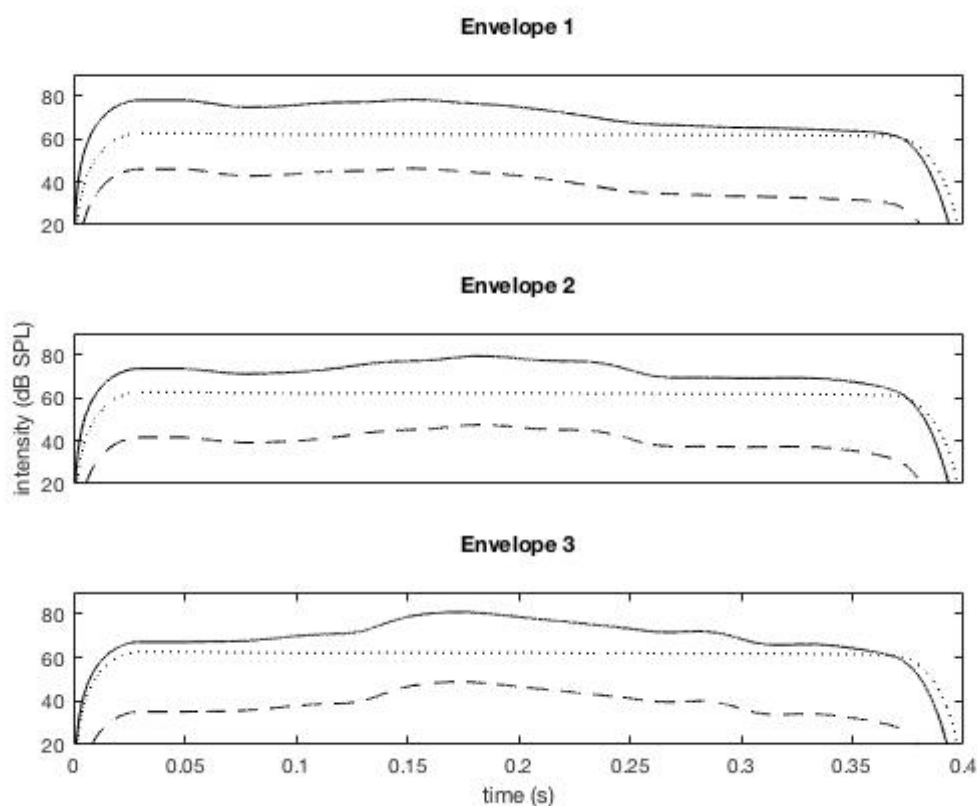


Figure 3-1 schematic representation of double-vowel pairs.

The dotted line represents the envelope of 1 female vowel. The solid line represents male vowels scaled to a RMS of 72 dB SPL (7-month-old infant stimuli), the dashed line represents male vowels scaled to a RMS of 40 dB SPL (adult stimuli).

Using Matlab (MATLAB 2014a, The MathWorks, Natick, 2014), the amplitude envelopes were extracted by computing the Hilbert envelope and low-pass filtering with a 2<sup>nd</sup> order butterworth filter with a 30-Hz cutoff frequency two times, forward and backward. A 30-ms raised-cosine onset and offset ramp was applied. Given that the male vowels in the baseline condition had some natural level variation, the central third of each baseline vowel was selected and elongated to 400 ms using the PSOLA algorithm in PRAAT. These vowel tokens were then multiplied by

the three different amplitude envelopes, resulting in 21 male vowels for the envelope cue condition. The average level of these vowels was scaled to 40, 72, and 75 dB SPL to produce TMRs of -20, +12 and +15 dB, respectively. Vowels were combined as in the baseline condition, resulting in a total of 36 vowel pairs at each TMR and for each amplitude envelope, 30 without a target vowel and 6 with a target vowel

### *Procedure*

All listeners were tested using the observer-based procedure (Werner, 1995), which allows for the inclusion and direct comparison of infants and adults. Testing was conducted in a sound attenuated booth. The sound booth contained a visual reinforcement tower, consisting of 2 boxes that concealed mechanical toys behind dark plexiglass and a small video monitor to show short video clips. All auditory stimuli were presented from a loudspeaker (JBL, Professional Control 1 or FOSTEX, Digital Personal Monitor 6301D) located at the front of the sound booth.

During testing, the infant was seated on the caregiver's lap. The reinforcement tower was two to the infant's right. An assistant was seated to the infant's left. The assistant manipulated quiet toys to keep the infant engaged and facing forward. An observer was seated outside the sound booth and watched the infant through an observation window. The caregiver, assistant and observer listened to double-vowel pairs presented without inter-stimulus-interval over insert earphones or circumaural headphones to prevent them from hearing the sounds in the sound booth.

The vowel pairs were played throughout the test session with a 1200-ms interstimulus interval. When the infant was facing forward and appeared attentive, the observer initiated a trial. During a trial, the infant either heard a vowel pair containing the target vowel (target trial) or a vowel pair that didn't contain the target vowel (no-target trial). From trial onset, the observer had

4 s to judge whether or not a target trial had occurred. The observer was blind to trial type. Observer judgments were based solely on the infant's behavioral response. Some typical infant responses indicating target trials included head turns towards the reinforcement tower, eye widening, or brief cessation of motor activity. The observer received feedback after each trial. When the observer correctly identified a target trial, a mechanical toy or a video was activated as visual reinforcement. Additionally, the assistant in the sound booth provided "social reinforcement" by clapping, smiling, and verbally praising the infant. When the observer incorrectly identified a target trial or correctly identified a no-target trial, no reinforcement was provided.

All listeners completed 4 test phases: a demonstration phase, a training phase, a baseline test phase and an envelope cue test phase. Target vowel and the order of test phase presentation were counterbalanced across listeners.

The demonstration phase served to introduce the target vowel and the task. The vowels in a pair had baseline amplitude envelopes. Vowel pairs with no target were presented at 0 dB TMR; target vowel pairs were presented at +15 dB TMR. A maximum of 20 trials were presented in the demonstration phase, with target and no-target trials occurring at a ratio of 5:1. The visual reinforcer was activated 2 s after target-trial onset, regardless of the observer's response. The demonstration phase was complete when the infant-observer team correctly identified 4 of the last 5 target trials and 1 no-target trial. The demonstration phase was discontinued if the infant's state precluded further testing or when 20 trials had been presented. The infant only continued to the second phase if the demonstration phase was completed.

The second phase, the training phase, served to teach the infant that a behavioral response was required for reinforcer activation and to familiarize the observer with the infant's responses.

The vowels in each pair had baseline amplitude envelopes and were presented at +15 dB TMR. Target and no-target trials occurred with equal probability. Reinforcement was only provided when the infant-observer team correctly identified a target trial. The training phase was complete when the observer correctly identified 4 of the last 5 target and 4 of the last 5 no-target trials. The training phase was discontinued if the infant's state precluded further testing or until a maximum of 40 trials had been presented.

Infants who completed the training phase completed two test phases: one in the baseline condition and one in the envelope cue condition. Each test phase consisted of 36 trials: 15 target trials, 15 no-target trials and 6 probe trials. On probe trials, the target-vowel pair was presented at +15 dB TMR, as a check that the infant was still engaged in the task. All other vowel pairs were presented at a +12 dB TMR. Trials were presented in 2 blocks: one block consisted of 7 target and 7 no-target trials, the other of 8 target and 8 no-target trials<sup>4</sup>. The order of block presentation was counterbalanced across listeners. Each test phase was continued until all 36 trials were completed, until infant state precluded further testing, or until an infant failed to

---

<sup>4</sup>In the baseline condition, the block with 7 target vowel pairs used the target vowel with the competing vowels /æ, ʊ, ɔ:, ɪ, æ, ʊ, and i: or u: (whichever vowel was not the target vowel)/. The block with 8 target vowel pairs used the target vowel with the competing vowels /ε, ɪ, æ, ʊ, ɔ:, ε, ɪ, and i: or u:(whichever vowel was not the target vowel)/. In the cue conditions, the envelopes were also preselected. The block with 7 target vowel pairs contained target+envelope 1 with the competing vowels /æ, ɪ/, target+envelope 2 with the competing vowels /ʊ, and i: or u: (whichever vowel was not the target vowel)/, and target+envelope 3 with the competing vowels /æ, ʊ, ɔ:/. The block containing 8 target vowel pairs presented target+envelope 1 with the competing vowels / ʊ, ɔ:, i: or u:(whichever vowel was not the target vowel)/, target+envelope 2 with the competing vowels /ε, ɪ, æ/, and target+envelope 3 with the competing vowels /ε, ɪ/.

respond to 3 probe trials. The test phase in one condition had to be completed before beginning the test phase in the second condition.

Infants completed all testing in 2-3 visits to the lab. Generally, the test phases for the two experimental conditions were completed on 2 separate visits to the lab. If testing was discontinued before reaching the training criterion, the next visit started with the demonstration phase. If testing was discontinued before completing a given test phase, the next visit began with a few reminder trials from the demonstration and training phases.

Adults were tested with the same procedure with some age-related adjustments. Adult listeners were seated in the sound booth by themselves and were instructed to raise their hand whenever they heard the sound “that makes the toys come on”. The observer watched through the observation window and recorded the responses. Adults completed the same phases infants completed, but TMRs were adjusted to prevent ceiling performance: The TMR in the training phase and on probe trials in the test phase was -10 dB, the TMR in the test phases was -20 dB. Adults completed all testing in one 1-hour visit to the lab with a short break between the two test phases.

## ***B. Results***

The measure of performance was  $d'$ . Perfect hit and false-alarm rates were adjusted by  $1/n$ , where  $n$  is the number of trials.

Figure 2 depicts individual  $d'$  scores in the baseline and envelope cue conditions. As intended, most adults'  $d'$  was approximately 1 in the baseline condition. In contrast, most infants'  $d'$  was below 1, in fact below chance performance, in the baseline condition. Variability in baseline performance is not too surprising given that listeners were tested with a fixed TMR that may not have been appropriate for every individual listener. Nonetheless, 9 out of 12 infants and

10 out of 13 adults appear to perform better in the envelope cue condition than in the baseline condition.

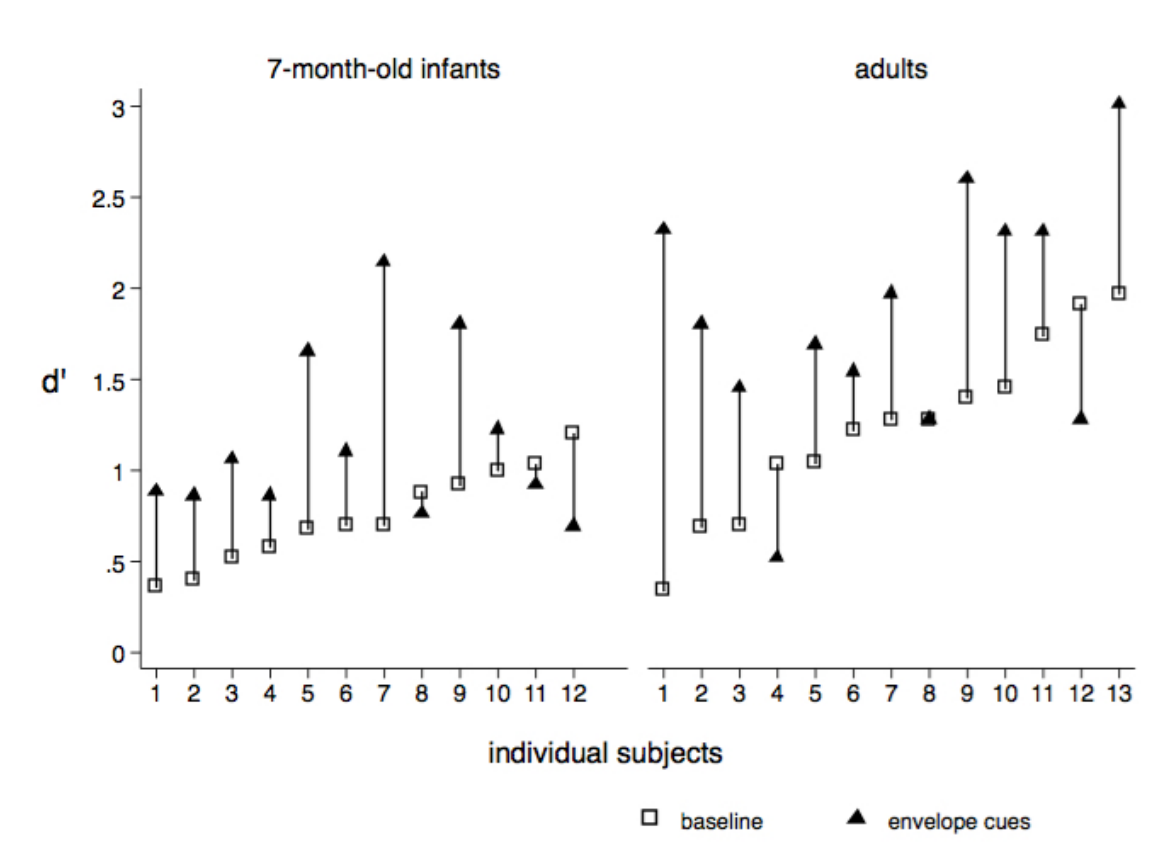


Figure 3-2 Individual  $d'$  scores in the baseline and envelope cue condition.

Individual subjects are listed along the x-axis, ordered by level of performance in the baseline condition. Squares represent performance in the baseline condition, triangles represent performance in the envelope cue condition. Each subject's  $d'$  scores in the two conditions are connected by a solid line.

Descriptive statistics suggested target vowel and order of test block presentation did not affect  $d'$ . While the descriptive statistics suggested that order of condition (baseline, envelope cue) may have interacted with age and cue condition, including it in the statistical model did not change the results. Thus, results were collapsed over target vowel, test block order and condition order.

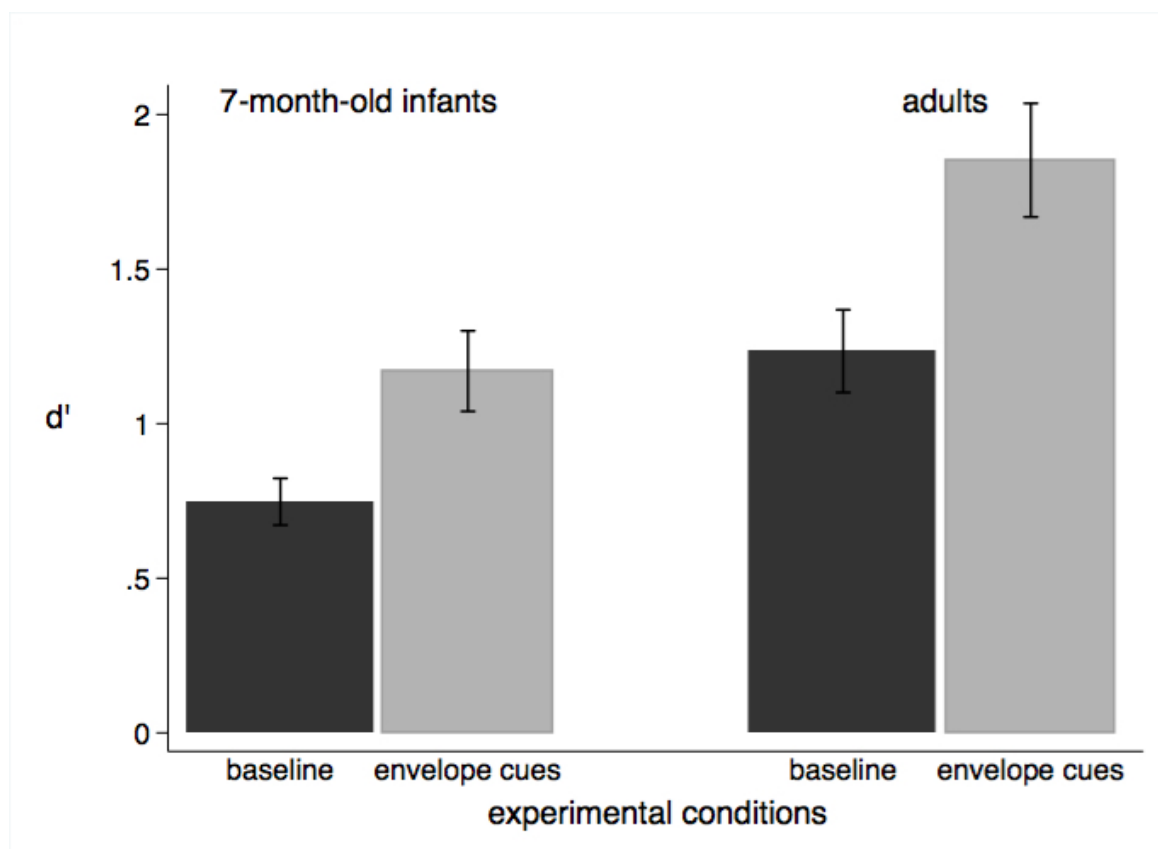


Figure 3-3 Average  $d'$  in the baseline and envelope cue conditions for infants and adults. Error bars represent +/- 1 standard error of the mean.

Figure 3 shows average  $d'$  in the baseline and envelope cue condition for both age groups. It is apparent that average  $d'$  in both conditions is lower for infants than for adults. However, both infants and adults seem to perform better in the envelope cue than in the baseline condition.

To statistically verify these observations, a mixed models analysis using Generalized Estimating Equations (GEE) assuming an exchangeable correlation structure and robust standard errors was employed. The model included the dependent variable  $d'$  and the independent variables cue (baseline condition, envelope cue condition), age (7-month-old infant, adult), and their interaction. The results are summarized in Table 2. Only the main effects of age and cue reached statistical significance. Controlling for cue, average  $d'$  was 0.48 lower in infants than in

adults. Controlling for age, average  $d'$  was 0.62 higher in the envelope cue condition than in the baseline condition. That the cue X age interaction did not reach significance leads us to conclude that the benefit of the envelope cue was no different for infants than for adults.

Table 3-2 Results of mixed models analysis: the effect of age and envelope cue on  $d'$ .

Predictor	B	SE B (robust)	z	p	95% conf. Interval
age	-0.48	0.15	-3.22	0.001	-0.78 – -0.19
cue	0.62	0.19	3.21	0.001	0.24 – 0.99
age x cue	-0.2	0.24	-0.81	0.418	-0.67 – 0.28

### ***C. Discussion***

The results of Experiment 1 show that while 7-month-old infants, on average, had lower  $d'$  than adults in both experimental conditions, the improvement in performance associated with differences in envelopes was similar in the two groups.

Considering infants' immature AM sensitivity, these results are somewhat surprising. It is possible that infants' benefit appears adultlike only because performance in the baseline condition is so poor for most infants. If that were the case, a correlation between  $d'$  in the two conditions would be expected. However, that was not the case [ $r(10) = 0.0008, p = 0.998$ ]. Because infants in a previous study (Oster & Werner, under review), as well as pilot subjects, achieved a  $d'$  of about 1 in the same baseline condition used here, it is likely that sampling error is responsible for the poor baseline performance observed in this study.

That infants receive adultlike benefit of an envelope cue in sound source segregation suggests that their poor AM sensitivity is not the result of imprecise representation of the amplitude envelope in the auditory pathway. Werner (2013) came to the same conclusion based on the fact that increasing modulation depth of a modulated masker did not improve infants'

masked vowel discrimination performance. Also consistent with that conclusion is the fact that infants' as young as 3 months of age appear to be able to use amplitude envelope information to discriminate between consonants in vocoded speech (Cabrera et al., 2013; Cabrera & Werner, 2017). Infants' poor performance in detecting envelope cue in noise may well result from task-related nonsensory factors.

That adults in the current study gained a benefit from the envelope cue appears to conflict with the results of Summerfield et al. (1992). Using similar vowel pairs to the ones used here, Summerfield and colleagues reported that listeners' thresholds for detecting a target vowel did not improve as a result of imposing different sinusoidal AM rates on the two competing vowels. However, Summerfield et al. did not test the condition used here, in which the target was modulated while the masker was unmodulated. They did report a benefit of coherent frequency modulation when the target was modulated in an unmodulated masker, even though no benefit was observed when target and masker were both modulated at different frequency modulation rates. A similar situation may hold for amplitude envelopes, as reported here. Clearly, this issue requires further systematic investigation.

### III. EXPERIMENT 2

Since the first year of life is marked by rapid auditory development (Eggermont & Moore, 2012; Moore & Linthicum, 2007), it might be expected that younger infants do not use envelope cues to the same extent as 7-month-old infants and adults. To address this question, 3-month-old infants' use of the same envelope cues used in Experiment 1 was investigated in Experiment 2. Ideally, 3-month-old infants would be tested using the same procedures as 7-month-old infants and adults to evaluate age effects in sensitivity. Although many 3-month-old infants were able to learn the task, most were unable to complete enough trials to obtain  $d'$  in two experimental

conditions. Thus, to gain some insight into 3-month-olds' use of envelope cues, we asked whether more infants can learn the vowel "identification" task with envelope cues than without. In addition, we evaluated whether 3-month-olds require fewer trials to learn the task with than without envelope cues. This approach is similar to that used in other developmental studies (e.g., Cabrera & Werner, 2017; Kuhl, 1979; Polka & Bohn, 2003). Infants recruited for Experiment 2 were only tested in the envelope cue condition. Performance was compared to that of 3-month-olds tested in the baseline condition in a previous experiment (Oster & Werner, under review).

### ***A. Method***

#### *Participants*

A total of 22 3-month-old infants (mean age: 12 weeks, SD = 0.7) participated in the envelope cue condition. Inclusion criteria were identical to those in Experiment 1. One additional infant was excluded due to a failed tympanometry screening. The 22 3-month-old infants (mean age: 12 weeks, SD = 0.7), who provided data in our previous experiment, met the same inclusion criteria (Oster & Werner, under review).

#### *Stimuli*

The stimuli used in Experiment 2 were the envelope cue vowel pairs used in Experiment 1, presented at the TMRs used in the demonstration and training phases of Experiment 1.

#### *Procedure*

Infants participated in the demonstration and training phases described in Experiment 1, using the same passing criteria. Testing was completed in 1 or 2 visits to the lab. During each visit, testing was continued until the infant reached the 80% correct criterion in the training phase, until infant state precluded further testing, or until the maximum number of trials had

been reached. If a session had to be discontinued prior to reaching criterion, a second attempt was made during the same visit. Infants were only judged to have failed to reach criterion if they were unsuccessful in at least 2 attempts in each of 2 visits.

## ***B. Results***

Two measures of task difficulty were considered 1) the proportion of infants reaching criterion with and without the envelope cue, and 2) the number of trials required to reach criterion with and without the envelope cue.

### *Percentage of infants reaching criterion*

In the baseline condition, 14 of 22 (64%) infants reached criterion. In the envelope cue condition, 16 of 22 (73%) infants reached criterion.

Two analyses were performed to evaluate whether the proportion of infants who successfully reached criterion was greater than would be expected by chance and whether the proportion was greater in the envelope cue condition than the baseline condition.

The first analysis simulated the proportion of infants who would reach the 80% correct criterion by chance. The simulation was based on 10,000 infants and assumed that the observer guessed on every trial. The same rules as the test procedures were used for each test phase: target probability, maximum number of trials, and maximum number of attempted sessions. From data in both conditions, the observer's response rate (# responses/# trials) in each phase was calculated. A "response" was defined as the observer having judged that a target trial had occurred, regardless of the trial type. The response rate in the demonstration phase was 0.68, in the test phase it was 0.58. The probability of aborted test sessions due to infant state was also included in the simulation. In the 1<sup>st</sup> session, the proportion of aborted sessions was 0.13, in the 2<sup>nd</sup> session 0.08, in the 3<sup>rd</sup> session 0.05, and in the 4<sup>th</sup> session 0. Based on this simulation, 44% of

infants would be expected to reach criterion by chance. One-sided exact binomial tests showed that the proportion of infants reaching criterion was greater than expected by chance in the envelope cue condition ( $p = 0.006$ ), but just missed statistical significance in the baseline condition. ( $p = 0.051$ ).

The second analysis evaluated whether the proportion of infants reaching criterion with the envelope cue was greater than the proportion of infants reaching criterion in the baseline condition. A one-tailed statistical test of proportions showed that the proportions of infants reaching criterion did not differ between conditions ( $z = -0.65, p = 0.259$ ).

#### *Number of trials required to reach criterion*

The number of trials required to reach criterion in the training phase was determined for each infant who reached criterion. To capture stimulus exposure, all trials, including those obtained from unsuccessful sessions, were included. Figure 4 shows a box-and-whisker plot of number of trials required to reach criterion in both conditions. A two-sided  $t$ -test showed that the number of trials infants required to reach criterion did not differ between the envelope cue and baseline condition [ $t(28) = -0.29, p = 0.747$ ].

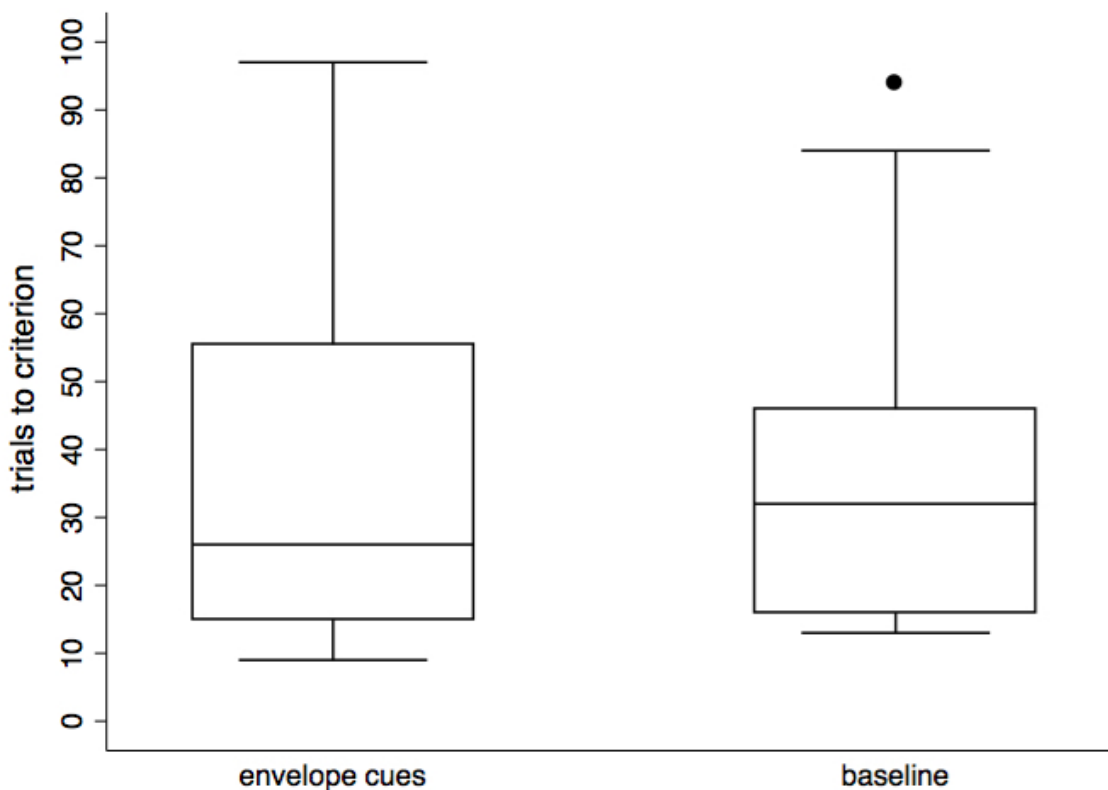


Figure 3-4 Box-and-whisker plot of trials-to-criterion as a function of cue condition.

The median number of trials is indicated by the horizontal lines inside the boxes. The upper and lower edges of each box represent the 75th and 25th percentile, respectively. Whiskers represent the minimum and maximum. Dots represent outliers. In the envelope cue condition,  $n = 16$ ; in the baseline condition,  $n = 14$ .

### *C. Discussion*

Experiment 2 showed that the proportion of 3-month-old infants learning the experimental task did not differ between the baseline condition and the envelope cue condition. For those infants who successfully reached the training criterion, learning the task appeared to be no less difficult in the envelope cue than in the baseline condition, based on the number of trials required to reach criterion. In other words, we find no evidence that 3-month-olds take advantage of modulation to separate a target vowel from a masker vowel.

The current results are similar to those of previous studies of 3-month-old infants in the number of trials required to reach criterion (e.g., Cabrera & Werner, 2017; Lau, Lalonde, Oster, & Werner, 2017). Moreover, Oster & Werner (under review) found that a higher proportion of infants learned to respond to one vowel of a pair when the onset of the target was delayed relative to that of the masker than when the onsets of the target and masker were synchronous. Thus, it is not the case that a relatively small proportion of infants can learn this task in any condition. One possibility that remains to be explored is that 3-month-olds are distracted by AM, and that distraction counteracts any benefit of AM for sound source segregation. A similar argument has been made to account for the reduced benefit of masker modulation and the failure to find benefits of temporal cues in tone detection among 7-month-olds (Werner, 2013; Werner et al., 2009).

#### IV. GENERAL DISCUSSION

The experiments reported provide the first results bearing on infants' use of envelope cues in sound source segregation. Three-month-old infants show no evidence of using differences in amplitude envelopes to segregate competing vowels. In contrast, although previous studies suggest that AM is not a salient cue for infants, the current results show that 7-month-olds gain an adultlike benefit of target AM in a segregation task.

Given that 3- and 7-month-old infants show similar sensitivity to AM (Werner, 2006), the results from these experiments suggest that it is the ability to use differences in the amplitude envelopes in sound source segregation that matures over this age range. The neural circuitry directly subserving sound source segregation is certainly developing in the first six months (Eggermont & Moore, 2012; Gorga, Kaminski, Beauchaine, Jesteadt, & Neely, 1989; Ponton, Moore, & Eggermont, 1996). Another explanation is that infants have to learn to use envelope

cues. Infants may initially require more salient cues, such as onset asynchrony, to segregate sounds and through experience learn to use less salient correlated cues, such as envelope cue. A similar model has been offered in explanation of age-related changes in the cues used for speech discrimination (Mayo & Turk, 2005; Nittrouer & Lowenstein, 2014; Seidl & Cristia, 2008).

Seven-month-olds' AM processing has been variously characterized as dreadful (AM detection; Werner, 2006) to inefficient (vowel discrimination in a modulated noise masker; Werner, 2013) to adultlike (vowel segregation; current study). It is likely that these different tasks engage different mechanisms. It is also likely that infants' responses depend to some extent on the stimuli—tones, noises, speech—involved. Taken together, the results of the studies of AM processing by infants suggest that infants have access to envelope information, but that whether they use that information is highly dependent on the context.

A related consideration is whether TMR in the vowel pairs in the envelope cue condition could have varied sufficiently over the course of the stimulus to allow listeners to focus on periods of relatively advantageous TMR and thereby improve their performance. In other words, does AM as used here actually improve sound source segregation as opposed to allowing a dip-listening strategy? To address this question, the 3 amplitude envelopes of the target vowels, for infants and for adults, are plotted along with that of the masker vowels (Figure 1). Note that the TMR for Envelope 1 was about 2 dB higher than the average TMR during the first 200 ms of the vowel pair. Similarly, the TMR for Envelope 2 was about 3.5 dB higher and Envelope 3 about 4 dB higher than the average TMR in the middle of the vowel pair. Thus, one might predict that listeners would perform better with Envelopes 2 and 3 than with Envelope 1. However, no differences in performance among envelopes were found in either infants or adults. Thus,

transitory increases in TMR are likely not the primary mechanism leading to improved performance in the envelope cue condition.

A related issue is whether the envelope cue was functionally an onset asynchrony. If the TMR was low enough near the beginning of the vowel pair, the target-talker vowels may only have become audible after about 150 ms (e.g., Envelopes 2 and 3). However, two facts argue against that idea. First, if listeners had primarily relied on onset-asynchrony cues, they would be expected to show less evidence of segregation for vowels with Envelope 1, which introduced the least amount of onset asynchrony. This pattern was not observed in either age group. Second, 3-month-old infants identified target vowels better with onset-asynchrony cues than without (Oster & Werner, under review), but did not perform better with envelope cue cues than without (current study). Although this makes it unlikely that listeners relied on asynchrony rather than envelope cues, a systematic evaluation of this question is warranted.

To summarize, the results of this study suggests that the ability to use differences between the amplitude envelopes of competing speech sounds undergoes rapid development between 3 and 7 months of age. While 7-month-old infants do not segregate concurrent vowels as efficiently as adults in general, envelope cues improve segregation to the same extent in 7-month-olds and adults. Thus, by 7 months of age, immature envelope cue use does not appear to be a major contributor to infants' difficulties processing one sound in the presence of others. Based on the results of this study, it would be predicted that infants with CI will develop the ability to use envelope cues in sound segregation, as long as these cues are consistently available to them. This prediction should be tested in future research.

## CHAPTER 4 INFANTS' USE OF COMBINED ONSET ASYNCHRONY AND ENVELOPE CUES

### I. INTRODUCTION

Infants' ability to process speech in competing sounds is immature for reasons that are poorly understood (Werner, 2017). Research over the past decades has shown that auditory encoding is mature early in infancy and, thus, cannot explain infants' difficulties processing competing sounds (Buss et al., 2012). However, sound source segregation does not only rely on accurate encoding but also on similarities and differences in the acoustic characteristics between competing sounds. How infants use sound characteristics to segregate competing sounds has not been established.

The spectral, temporal or spatial location characteristics that support sound segregation in adults are well documented. For example, adults are better able to segregate sound sources with different fundamental frequencies or onsets (Bregman, 1990). Although existing studies have not systematically examined infants use of such auditory cues in sound source segregation, their results suggest that infants can use frequency separation and harmonicity (e.g., Bendixen et al., 2015; Demany, 1982; Folland et al., 2015; Folland et al., 2012; Winkler et al., 2003), spatial location (Nozza et al., 1990), onset asynchrony (Oster & Werner, under review), and envelope cues (Oster & Werner, in preparation). However, all of these studies investigated infants' use of auditory cues in isolation, a scenario that is only found in the laboratory. In natural situations, the components of sounds coming from different sources are likely to differ along multiple dimensions.

Importantly, the adult literature shows that not all cues that are simultaneously available contribute equally to sound source segregation. For example, Hall et al. (2006) examined the role

of onset asynchrony and envelope cues in the detection of a 2-kHz narrow-band noise masked by flanking narrow noise bands. Threshold for the target band depended on a complex interaction of envelope cues, onset asynchrony and variability in flanking band frequency. Moreover, the specific effects of these variables were different for different listeners. That age also influences thresholds in this situation was demonstrated by Hall et al. (2008). In that study, 4- to 11-year-old children were found to exhibit smaller threshold improvements than adults when an envelope cue was available, but about the same threshold improvement as adults when an onset asynchrony cue was available. Furthermore, while for children the effect of an onset cue was the same whether or not the envelope cue was also available, for adults the effect of an onset cue was reduced when an envelope cue was also available.

We have previously shown that 7-month-old infants were able to use onset asynchrony cues and envelope cues to segregate concurrent vowel pairs as well as adults (Oster & Werner, under review; Oster & Werner, in preparation). In addition, we found that 3-month-old infants were better at segregating concurrent vowel pairs with an onset asynchrony than without, but that they were no better at that task with an envelope cue than without. The current study asks how 7- and 3-month-old infants segregate concurrent vowel pairs when both onset asynchrony and envelope cues are available. While no previous studies have addressed this question directly, the results of several studies indicate that infants may use different cues from adults (Cabrera & Werner, 2017; Mayo & Turk, 2005; Nittrouer & Lowenstein, 2014) or that they require more cues than adults (Cooper & Aslin, 1994; Seidl & Cristia, 2008) to speech discrimination. Hall et al.'s (2008) results with older children also suggest that infants may be different from adults in the way they combine auditory cues in sound source segregation.

Two experiments were conducted to determine how infants use a combination of onset asynchrony and envelope cues. The first experiment measured the benefit of combined onset asynchrony and envelope cues in concurrent vowel segregation in 7-month-old infants and adults. The second experiment evaluated 3-month-old infants' use of these combined cues in concurrent vowel segregation. The results from both experiments were compared to those of our prior experiments to evaluate the contribution of the individual cues to the benefit gained from combined cues.

## II. EXPERIMENT 1

### *A. Method*

#### *Design*

Listeners were presented with a train of “double-vowel” pairs. One vowel in the pair was spoken by a man, the other by a woman. Listeners were trained to respond to one target vowel, either the male /i:/ or male /u:/. Experiment 1 employed a mixed model design to measure the benefit, in  $d'$ , of combined temporal cues to concurrent vowel segregation. Results were also compared to those obtained in 2 prior experiments that measured the benefit, in  $d'$ , of isolated onset asynchrony and isolated envelope cues. In all experiments the double-vowel pairs were presented at the same age specific target-to-masker ratios (TMRs), which had produced a  $d'$  of approximately 1 in the baseline condition in pilot testing. An increase in  $d'$  in the combined cue condition over the baseline condition was taken as evidence that listeners gained a benefit from combined temporal cues. The difference in the size of the benefit between combined and isolated temporal cues was taken as a measure of how much individual cues contributed to the combined benefit.

### *Participants*

Infants' families and adult participants were recruited through the Communication Studies Participant Pool at the University of Washington, flyers, and newspaper announcements.

A total of 13 7-month-old infants (mean age: 30 weeks, SD = 2.3) and 12 young adults (mean age: 21 years, SD = 2.3) completed testing in both experimental conditions. The inclusion criteria for infants included passing newborn hearing screening, normal development, no family history of congenital hearing loss, no risk factors for hearing loss, no treatment for otitis media within 2 weeks of testing, passing tympanometry screening with a 1000 Hz probe tone on each test date, English as the primary language in the home, and no prior experience with psychoacoustic experiments. Inclusion criteria for adults were identical except that adults were required to pass a hearing screening bilaterally at 20 dB HL at octave frequencies between 250 and 4000 Hz during the first visit, and tympanometry screening was performed with a 226 Hz probe tone.

An additional 6 participants were excluded for the following reasons: failed tympanometry screening (2 infants), below chance performance in both experimental conditions (1 infant and 1 adult), failed hearing screening (1 adult), and prior participation in psychoacoustic experiments (1 adult).

### *Stimuli*

The 7 American-English vowels /i: u: ʊ æ ε ɔ: ɪ/ were used to create double-vowel pairs. Vowels were recorded by a male and a female talker in a professional recording studio with a high-quality condenser microphone. A sampling rate of 44100 Hz was used and recordings were digitized at 32 bits. One exemplar of each vowel with relatively flat temporal envelope was chosen from both talkers. The vowels from each talker were further processed in PRAAT

(Boersma & Weenink, 2017). All vowels were shortened to 400 ms, including 15-ms long on- and off-ramps. For each talker, one vowel was chosen and the pitch contour extracted using the PSOLA algorithm in PRAAT. This pitch contour was applied to all vowels from that talker. Fundamental frequencies (F0) of the female vowels ranged between 219 and 235 Hz, F0s of the male vowels ranged between 83 and 123 Hz. Vowels were then scaled in root-mean-square (RMS) amplitude using the scale intensity function in PRAAT. Female vowels were scaled to 60 dB SPL, male vowels were scaled to 40, 50, 60, 72, and 75 dB SPL. The male talker was always the target talker, and the vowels /i:/ or /u:/ were the target vowels. Target vowels were selected because their first and second formants had minimal overlap with the first and second formants of the female vowels.

These vowels were used to create double-vowel pairs. Each vowel pair consisted of 2 different vowels, 1 from the male and 1 from the female talker. The target vowel was never used as a competing vowel; that is, if the target vowel was /i:/, then the female /i:/ was not included among the competing vowels. Thus, 7 male and 6 female vowels were used to generate a total of 36 vowel pairs, 6 of which contained a target vowel. Since the male vowels had 5 different intensities, double-vowel pairs with 5 different TMRs were created: -20, -10, 0, +12, and +15 dB.

For the combined cue condition, the male vowels were further processed before being combined with the female vowels. First, the male vowels were shortened by selecting the central 300 ms of each vowel token. This vowel segment was then multiplied by 3 different amplitude envelopes, which had been obtained from vowel tokens spoken by the same male talker in infant-directed speech. The modulated vowel tokens were shortened to 300 ms using the PSOLA algorithm in PRAAT. The amplitude envelopes were extracted in Matlab (MATLAB 2014a, The

MathWorks, Natick, 2014), by computing the Hilbert envelope and low-pass filtering with a 2<sup>nd</sup> order butterworth filter with a 30-Hz cutoff frequency, forward and backward. Then a 30-ms raised-cosine onset and offset ramp was applied and vowels were scaled to 40, 72 and 75 dB SPL. Last, zero padding was added to the beginning of the vowels to equate them in length to the female vowels. In this manner, 7 male vowels were created at each of three intensities and with each of the 3 male vowel envelopes.

The combined cue male vowels were combined offline with the female vowels in the same manner as described above. The result was a total of 90 vowel pairs without a target vowel and 18 vowel pairs with a target vowel at TMRs of -20, +12 and +15 dB. A schematic representation of combined cue vowel pairs at +12 dB TMR is depicted in Figure 1.

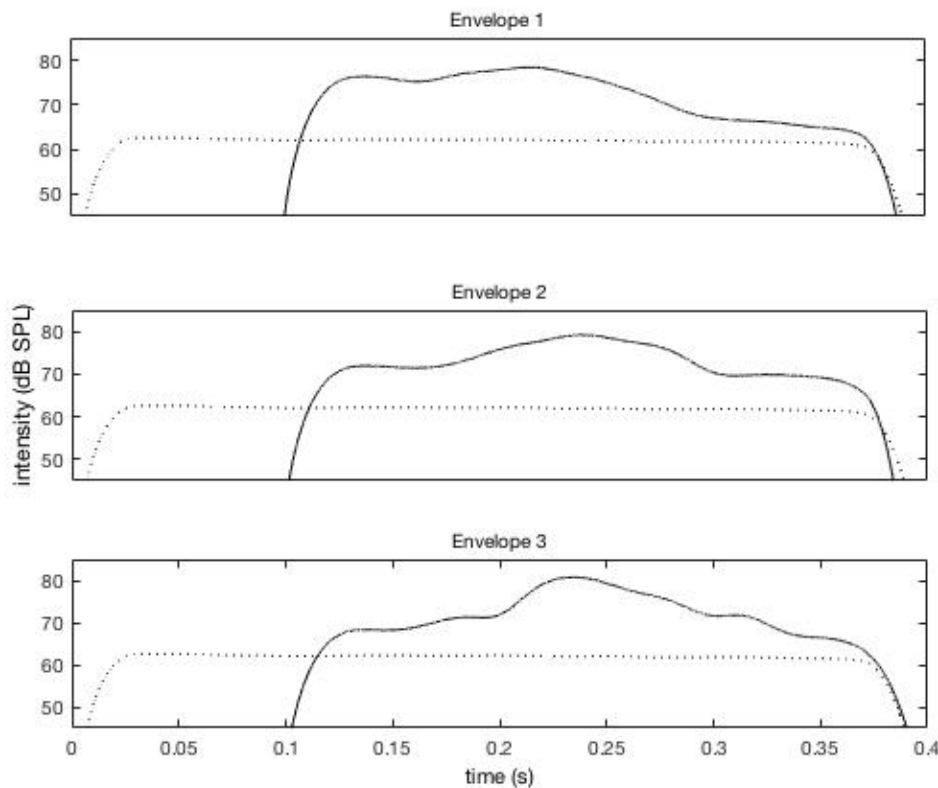


Figure 4-1 schematic representation of vowel pairs in the combined cue condition.

The dotted line represents one female vowel, the solid line one male vowel. The 3 different envelopes are represented in the 3 panels.

*Procedure*

All listeners were tested using the observer-based psychoacoustic procedure (Werner, 1995). During testing, the infant was seated on the caregiver's lap inside a sound-attenuated booth. In front and to the left of the infant was an assistant who manipulated quiet toys to keep the infant engaged. In front and to the right of the infant was a visual reinforcement tower. The visual reinforcement tower consisted of 2 stacked boxes containing a light bulb and a mechanical toy, concealed behind dark plexiglass, and a small TV monitor. A loudspeaker was located directly in front of the infant (JBL; Professional Control 1 or FOSTEX, Digital Personal Monitor 6301D). The caregiver and the assistant wore headphones and listened to masking sounds, consisting of double-vowel pairs presented with no silent interval between pairs. The observer was seated outside the sound booth and was able to view the infant, caregiver and assistant through an observation window. The observer listened to the same masking sounds over headphones and controlled the experiment via a custom computer program.

Throughout the test session, the infant was presented continuously with double-vowel pairs at a rate of 1 vowel pair every 1.6 seconds. Regardless of the test phase, the observer started a trial whenever the infant appeared ready. During a trial, the infant was either presented with a vowel pair containing a target vowel (target trial) or a vowel pair not containing a target vowel (no-target trial). The observer was blind to trial type and judged whether a target trial had been presented based on the infant's behavior. Typical infant responses taken as evidence of a target trial included head turns toward the reinforcer toys, eye widening, or brief cessation of motor activity.

The infant was introduced to the target vowel and experimental task during the demonstration phase. No-target vowel pairs were presented with a 0 dB TMR; target vowel pairs were presented with a +15 dB TMR. All vowel pairs had simultaneous onset and similar

amplitude envelopes. Target and no-target trials were presented at a ratio of 5:1. The observer had 2 s from trial onset to judge trial type. If the observer correctly identified a target trial, one of the visual reinforcers was activated, and the assistant provided social reinforcement. If the observer missed a target trial, the infant received reinforcement at the end of the 2 s observation window. The demonstration phase was completed if the infant-observer team correctly identified 4 out of the last 5 target trials and 1 no-target trial, but otherwise continued until infant state precluded further testing or until 20 trials had been presented.

After successful completion of the demonstration phase, infants completed the training phase. Here, all vowel pairs were presented with a +15 dB TMR, had simultaneous onset and similar temporal envelopes. The training phase was intended to teach the infant that a response was required to activate the visual reinforcer and to teach the observer what infant responses to target vowels might be. During a trial, target and no-target trials were equally likely. After trial onset, the observer had 4 s to judge trial type. The infant only received visual and social reinforcement when the observer correctly identified a target trial. When the observer's judgment was incorrect or when the observer correctly identified a no-target trial, no reinforcement was provided. The training phase was completed once the infant-observer team correctly identified 4 out of the last 5 target and 4 out of the last 5 no-target trials. If an infant did not reach criterion within 40 trials or if infant state precluded further testing, the session ended.

After successful completion of the training phase, the infant was presented with two test phases: a baseline test phase and an auditory cue test phase. Vowel pairs were presented with a +12 dB TMR. Target and no-target trials were equally likely to occur. In addition 6 probe trials, target trials with +15 dB TMR were presented. The purpose of the probe trials was to ensure that infants were still attending to the task. The observer had 4 s after trial onset to judge trial type,

and reinforcement was only provided on correctly identified target trials. The infant completed a total of 36 trials: 15 no-target, 15 target and 6 probe trials. Target, no-target and probe trials were grouped into 2 blocks, each containing 3 probe trials. One block contained 7 target and 7 no-target trials, the other block contained 8 target and no-target trials<sup>5</sup>. The target vowel pairs in each block were preselected to ensure that all listeners were presented with the same target-vowel pairs. The no-target vowel pairs were randomly chosen from the available no-target vowel pairs. The test phase continued until the infant completed all 36 trials, until an infant missed 3 probe trials or until infant state precluded further testing.

If testing was discontinued before the training criterion was reached, the next session began with the demonstration phase. Otherwise the next session started with a few reminder trials from the demonstration and training phase. Infants completed all testing in 2-3 visits to the lab.

Adults were tested in essentially the same way, except that they were seated in the sound booth by themselves and were instructed to raise their hand whenever they heard “the sound that makes the toy turn on”. Adults were tested at different TMRs than infants: -10 dB in the training phase and -20 dB in the testing phase. The age specific TMRs were chosen because they produced a  $d'$  of approximately 1 in the baseline condition in pilot testing and in previous studies. Adults completed all testing in 1 visit to the lab.

---

<sup>5</sup> In the baseline condition, the block with 7 target vowel pairs used the target vowel with the competing vowels /æ, ʊ, ɔ:, ɪ, æ, ʊ, and i: or u: (whichever vowel was not the target vowel)/. The block with 8 target vowel pairs used the target vowel with the competing vowels /ε, ɪ, æ, ʊ, ɔ:, ε, ɪ, and i: or u:(whichever vowel was not the target vowel)/. In the cue conditions, the envelopes were also preselected. The block with 7 target vowel pairs contained target+envelope 1 with the competing vowels /æ, ɪ/, target+envelope 2 with the competing vowels /ʊ, and i: or u: (whichever vowel was not the target vowel)/, and target+envelope 3 with the competing vowels /æ, ʊ, ɔ:/. The block containing 8 target vowel pairs presented target+envelope 1 with the competing vowels / ʊ, ɔ:, i: or u:(whichever vowel was not the target vowel)/, target+envelope 2 with the competing vowels /ε, ɪ, æ/, and target+envelope 3 with the competing vowels /ε, ɪ, /.

Target vowel, order of block presentation in the test phase, and order of experimental condition were counterbalanced. Note that 5 of the 6 adults who were tested in the baseline condition first provided data for a prior study before completing the combined cue condition. Data were only included in the final analysis if listeners performed above chance in at least one experimental condition. Assuming a binomial process with  $p = 0.5$ , the probability of a  $d' \geq 0.8$  with 30 trials would be less than 0.05 if an observer responded “yes” randomly on every trial.

### ***B. Results***

Performance was measured in  $d'$ . Cases of perfect hit and/or false-alarm rates were adjusted by  $1/n$ , where  $n$  is the number of trials. Descriptive statistics suggested that neither the target vowel nor the presentation orders of experimental blocks or condition affected  $d'$ . Consequently, results were collapsed over these categories for the following analyses.

#### *Combined onset asynchrony and envelope cues*

Figure 2 shows  $d'$  in the baseline and combined cue conditions for individual listeners. Performance is highly variable in both age groups and in both conditions. 8 of 13 infants and 7 of 12 adults performed better in the combined cue condition than in the baseline condition.

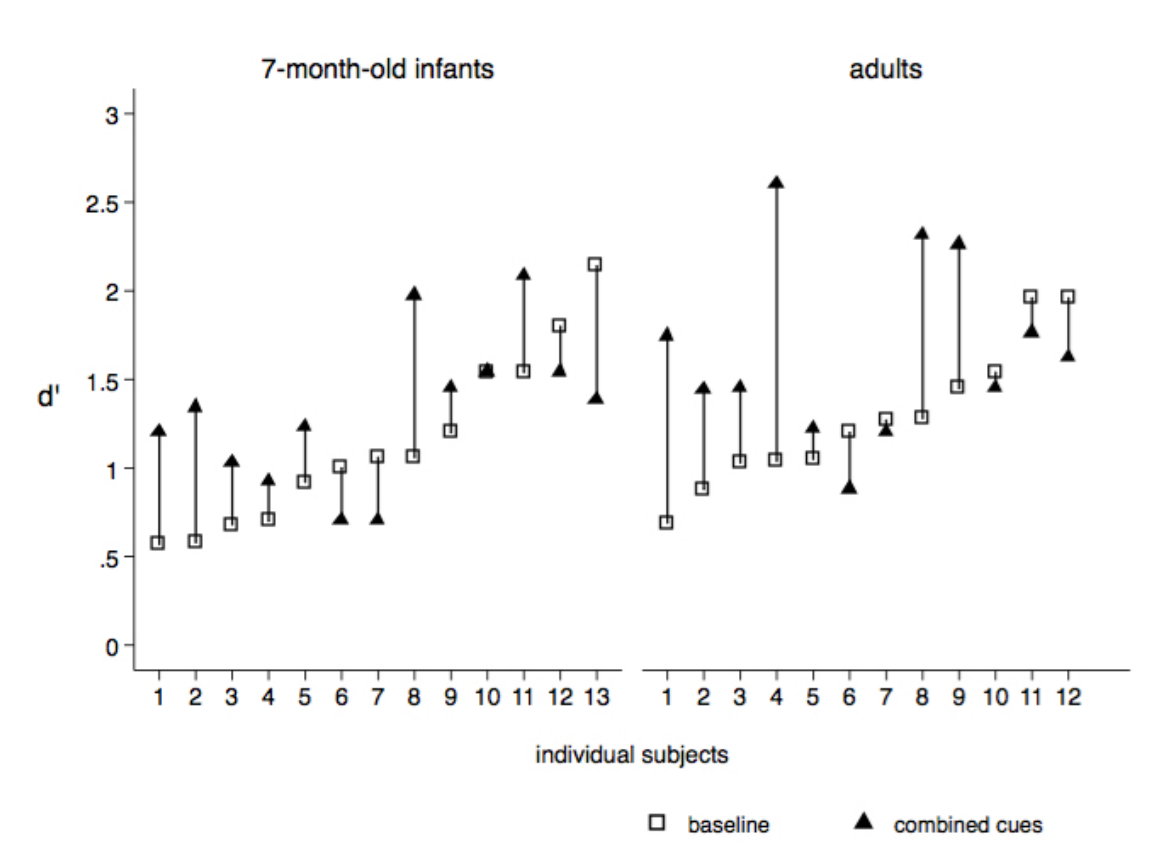


Figure 4-2 Individual  $d'$  scores in the baseline and combined cue conditions for 7-month-old infants and adults. Individual subjects are listed along the x-axis, ordered by baseline  $d'$ . Individual scores in the baseline condition (squares) and in the combined cue condition (triangles) are connected by a solid line.

Figure 3 depicts average  $d'$  in the baseline and combined cue conditions for 7-month-old infants and adults. As intended, performance in the baseline condition appears similar in infants and in adults. On average, both age groups perform better in the combined cue than in the baseline condition.

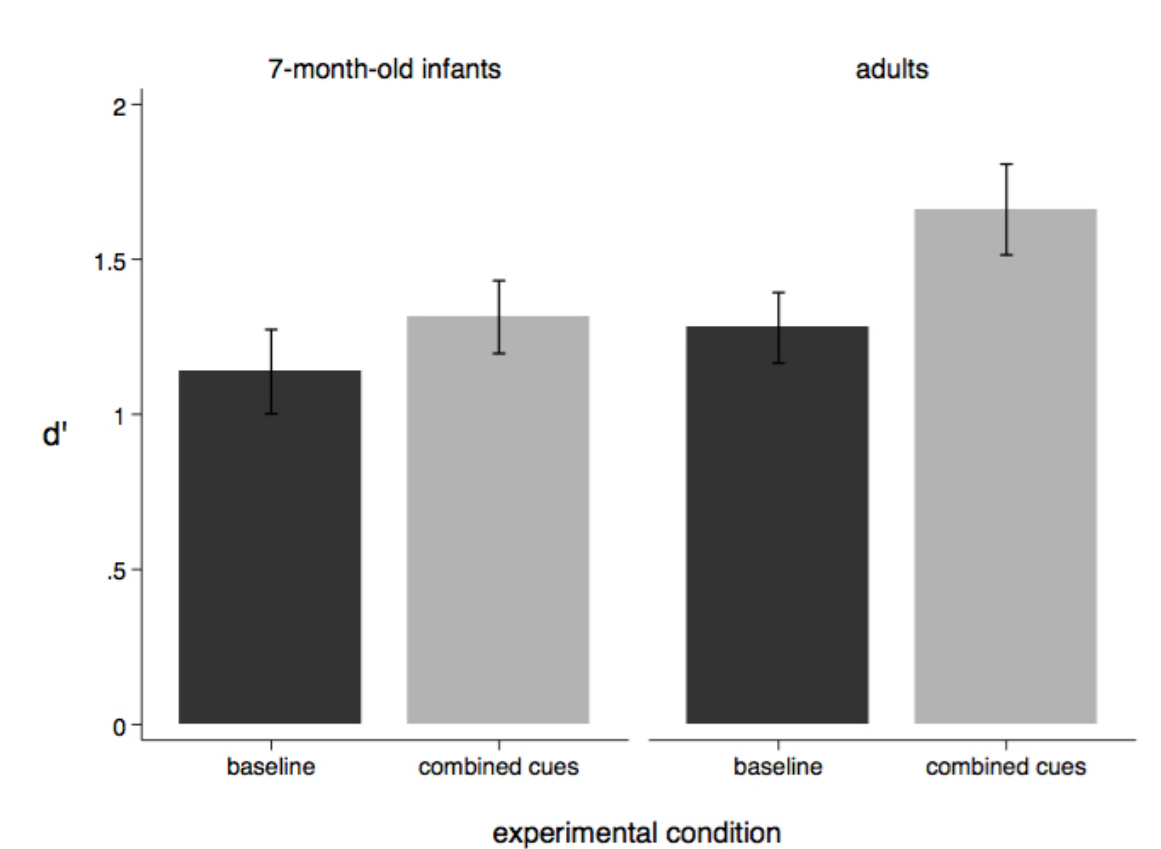


Figure 4-3 Average  $d'$  in the baseline and the combined cue conditions for infants and adults. Error bars represent +/- 1 standard error of the mean.

To verify these observations, a mixed models analysis was performed employing Generalized Estimating Equations assuming an exchangeable correlation structure and robust standard errors. The final model included the dependent variable  $d'$  and the independent variables age (infants, adults), cue (presence or absence), and their interaction. Table 1 summarizes the results. The only effect reaching statistical significance was cue: Regardless of age, average  $d'$  was 0.38 higher in the combined cue condition than in the baseline condition. The age X cue interaction failed to reach statistical significance, indicating that adults and infants gained a similar benefit from the addition of combined asynchrony and envelope cues.

Table 4-1 Results of mixed models analysis: the effect of age and combined cues on  $d'$ 

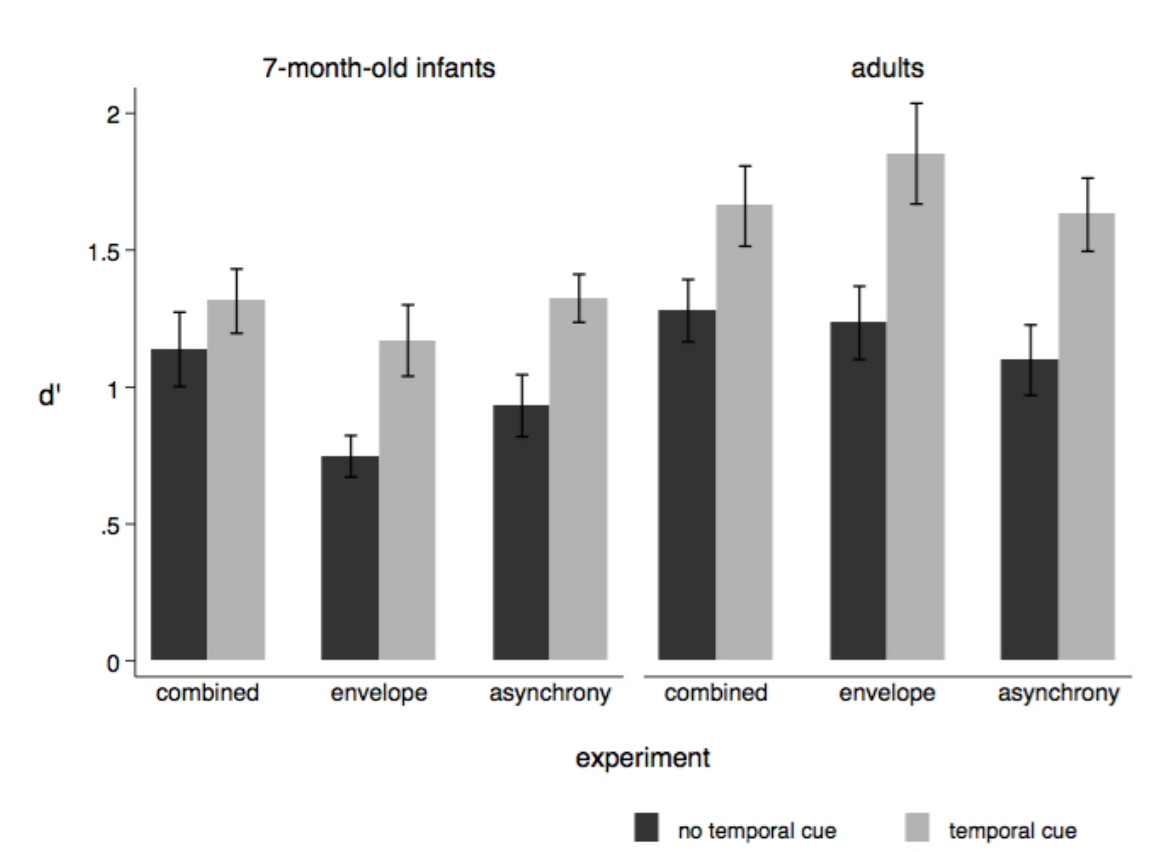
Predictor	B	SE B (robust)	z	p	95% conf. Interval
age	-0.14	0.17	-0.81	0.415	-0.48 – 0.2
cue	0.38	0.18	2.17	0.03	0.04 – 0.73
age x cue	-0.21	0.22	-0.93	0.352	-0.64 – 0.23

### *Combined cues versus isolated cues*

Two prior experiments had been conducted using identical procedures and the same vowel stimuli. In one a 100-ms onset-asynchrony cue was introduced (Oster & Werner, under review); in the other, the envelope cue used here was introduced (Oster & Werner, in preparation). Performance in those experiments was compared to that in the present experiment to answer the following questions: 1) Does performance improve more when onset asynchrony and envelope cues are combined compared to presenting either cue in isolation? and 2) If such an effect exists, does it depend on age? A mixed models analysis using Generalized Estimating Equations assuming an exchangeable correlation structure and robust standard errors was performed. The model included the dependent variable  $d'$  and the independent variables age (infants, adults), cue (presence or absence), experiment (combined cues, envelope cues, onset-asynchrony cues), and all possible interactions. The results of this analysis are summarized in Table 2. Only the main effect of cue reached statistical significance: Average  $d'$  was higher in the cue condition than the baseline condition, regardless of cue or age. These results are graphically summarized in Figure 4. The combined cues led to no better performance than either the asynchrony or the envelope cue.

Table 4-2 Results of mixed models analysis: the effect of age, presence and type of cue on  $d'$ .

Predictor	B	SE B (robust)	z	p	95% conf. Interval
age	-0.27	0.16	-1.66	0.098	-0.58 – -0.05
cue	0.44	0.17	2.58	0.01	0.11 – 0.78
experiment	-0.1	0.08	-1.21	0.228	-0.27 - 0.06
ageXcue	-0.22	0.21	-1.05	0.296	-0.63 - 0.2
experimentXage	-0.02	0.12	0.18	0.859	-0.22 - 0.26
experimentXcue	0.01	0.12	0.48	0.628	-0.18 - 0.3
experimentXcueXage	0.04	0.15	0.25	0.804	-0.26 - 0.34

Figure 4-4 Average  $d'$  for infants and adults in three experiments.

Black bars represent baseline performance; grey bars represent performance with temporal cues. Error bars represent +/- 1 standard error of the mean.

### *C. Discussion*

The results of Experiment 1 show that infants and adults gain a similar benefit from combined temporal cues in concurrent vowel segregation. This benefit was no greater than that gained from isolated onset asynchrony or envelope cues.

Both the infants' and the adults' results parallel those reported by Hall et al. (2006). Recall that in that study adults detected a narrowband noise centered at 2 kHz in the presence of flanking masker bands with variable frequency, a situation similar to competing speech. Conditions in which an envelope cue, an onset asynchrony cue, and a combination of envelope and asynchrony cues were tested. Hall et al. found that combining envelope cues and onset asynchrony cues did not improve target detection over that observed with onset asynchrony cues alone. As in that study, the effects of both isolated and combined cues varied across listeners. Although in a subsequent study, Hall et al. (2008) found that both 4- to 11-year-olds and adults performed better with a combination of envelope and asynchrony cues than they did with either cue alone, the frequencies of the flanking bands in that study did not vary from trial-to-trial. In that condition, adults also appear to use both asynchrony and envelope cues in the combined cue condition.

The results are not consistent with the prediction that additional cues would be particularly beneficial for infants' segregation. Is it possible that infants in the current study did not gain more benefit from combined cues than from isolated cues because their performance with isolated cues was at ceiling? We know that infants, on average, do not achieve perfect performance, even in the easiest conditions, due to inattentiveness. While infants' asymptotic performance on this particular task has not been established, if infants were as inattentive as previously reported in tone or noise detection, with asymptotic performance averaging 85%

correct (e.g., Bargones et al., 1995), they should be able to achieve an average  $d'$  of 2.1 in the easiest conditions. Average infant performance with isolated and combined cues was well below that value, leading to the conclusion that ceiling effects cannot account for infants' failure to perform better with combined cues than with isolated cues.

The effects of combined envelope and onset asynchrony cues observed here are no different in adults and 7-month-old infants. However, it is still possible that the individual cues contribute differently to the combined benefit in infants and adults. For example, infants could be using the envelope cue, while adults use the asynchrony cue. However, since previous studies have shown that onset asynchrony is a particularly strong cue for both adults and children (Hall et al., 2006; Hall et al., 2008), it is more likely that onset asynchrony is driving the combined cue benefit in both age groups.

### III. EXPERIMENT 2

Experiment 2 evaluated how 3-month-old infants use combined cues in concurrent vowel segregation. Younger infants appear to require more cues than older infants in speech perception tasks (Cooper & Aslin, 1994; Seidl & Cristia, 2008). Thus, 3-month-old infants might be expected to segregate competing vowels better with combined cues compared to the individual cues. However, our prior experiments showed that 3-month-old infants' vowel segregation improved with onset asynchrony cues but not with envelope cues (Oster & Werner, under review; Oster & Werner, in preparation). The most straightforward prediction, then, would be that 3-month-olds will benefit from the combined cues to the same extent as from onset asynchrony cues alone. In previous experiments, 3-month-old infants were unable to complete enough trials to obtain the same sensitivity measures used in Experiment 1. Therefore, Experiment 2 evaluated two variables reflecting the ease with which 3-month-olds learn to

respond to a target vowel in a vowel pair when both onset asynchrony and envelope cues are available. First, the proportion of infants reaching the 80% correct criterion in the training phase was assessed. Additionally, the number of trials required to reach the 80% correct criterion was used as another measure of task difficulty. Both measures have been used successfully in other developmental studies with young infants (e.g., Cabrera & Werner, 2017; Kuhl, 1979; Lau et al., 2017; Polka & Bohn, 2003). Given that these measures differ from those in Experiment 1, they do not allow for a direct comparison of 3-month-olds, 7-month-olds and adults.

### ***A. Method***

#### *Participants*

A total of 23 3-month-old infants (mean age: 12 weeks, SD = 0.9) provided data in Experiment 2. The inclusion criteria were identical to experiment 1. An additional 22 3-month-old infants (mean age: 12 weeks, SD = 0.7) provided data in the baseline condition in a previous experiments (Oster & Werner, under review). The inclusion criteria were identical to those reported here.

#### *Stimuli and Procedures*

The passing criterion and stimuli in the demonstration phase were identical to those in Experiment 1. While the passing criterion in the training phase remained identical to that of Experiment 1, the stimuli were pairs with the combined onset asynchrony and envelope cues, rather than baseline pairs. In other words, the combined cue stimuli from the test phase in Experiment 1 were presented in the training phase in Experiment 2. As in Experiment 1, the TMR in the training phase was +15 dB, and the maximum number of trials was 40.

Once an infant met criterion in the demonstration phase, testing continued in the training phase. If a session had to be discontinued due to infant state before the maximum number of

trials or 80% correct criterion was reached, a second session was attempted during the same visit if the infant appeared able to continue. Infants were only judged to be unable to reach the training criterion if it was not reached within 2 sessions in 2 visits to the lab.

## **B. Results**

### *Percentage of infants reaching criterion with combined cues*

The number of infants reaching criterion was 14 out of 22 (64%) in the baseline condition and 18 out of 23 (78%) in the combined cue condition.

Two analyses were performed to evaluate 1) whether more infants reached criterion with combined temporal cues than would be expected by chance, and 2) whether more infants reached the criterion in the combined cue condition than in the baseline condition.

The first analysis simulated the proportion of 10,000 infants who would reach training criterion if the observer guessed on every trial in the baseline and combined cue condition. The simulation used the same target probability, maximum number of trials and maximum number of attempted sessions as the test procedure in each test phase. Additionally, the observer's response rate in each phase was included in the model. This response rate was calculated from the data by dividing the number of times the observer judged that a trial had occurred, regardless of the trial type, by the total number of trials. The response rate in the demonstration phase was 0.67, in the test phase it was 0.6. The probability of aborted test sessions due to infant state was also included in the simulation. In the 1<sup>st</sup> session, the proportion of aborted sessions was 0.14, in the 2<sup>nd</sup> session 0.09, in the 3<sup>rd</sup> session 0.05, and in the 4<sup>th</sup> session 0. Based on this simulation, 43% of infants would be expected to reach criterion by chance. One-sided exact binomial tests showed that the proportion of infants reaching criterion was greater than expected by chance in both experimental conditions (baseline:  $p = 0.0416$ , combined cue:  $p = 0.0006$ ).

The second analysis, using a one-tailed statistical test of proportions, showed that the proportion of infants reaching criterion in the combined cue condition was not significantly greater than the proportion in the baseline condition ( $z = -1.04, p = 0.15$ ).

*Number of trials required to reach criterion*

A secondary analysis evaluated whether the number of trials required to reach criterion differed between experimental conditions. This approach provides an alternative measure of task difficulty: presumably more trials are required to learn a more difficult task.

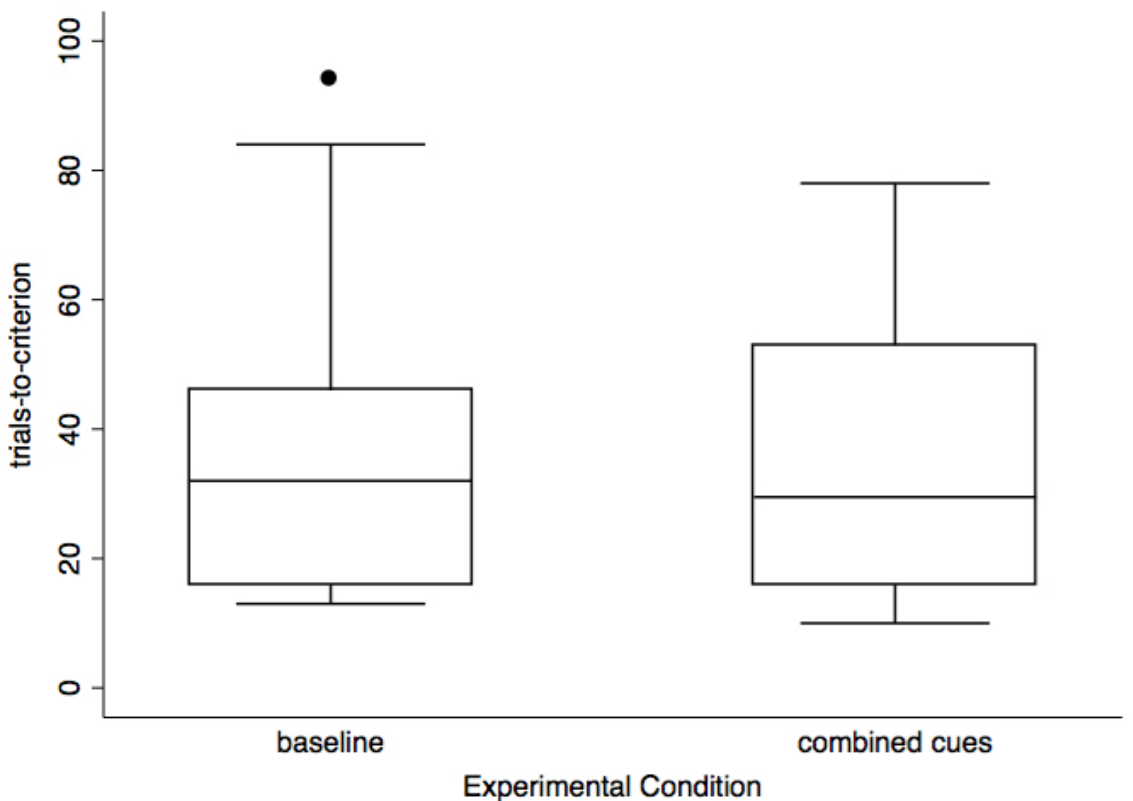


Figure 4-5 Box-and-whisker plot of trials to criterion with and without combined cues.

The median number of trials required to reach criterion is indicated by the horizontal line inside boxes. Lower and upper edges of the box represent the 25th and 75th percentile, respectively.

Whiskers represent the minimum and maximum number of trials required to reach criterion, and dots represent outliers.

As can be seen in Figure 5, the number of trials to reach the 80% correct criterion was similar between conditions. This observation was confirmed by a two-sided t-test [ $t(30) = 0.75$ ,  $p = 0.4586$ ]. This indicates that for infants who were able to reach the criterion, the task was similarly difficult in baseline and combined cue conditions.

### *C. Discussion*

Contrary to expectation, the results of Experiment 2 do not provide any evidence that combined onset asynchrony and envelope cues support concurrent vowel segregation in 3-month-old infants. Interestingly, for those infants who reached criterion, learning the task appeared similarly difficult with and without the two temporal cues.

Since our previous studies demonstrated that 3-month-old infants were better able to segregate concurrent vowels with onset asynchrony but not with envelope cues, these results do not support the notion that 3-month-old infants segregate competing speech sounds better with more cues. In fact, more infants appear to learn the task with only onset asynchrony cues than with asynchrony and envelope cues combined. The availability of envelope cues appears to have a detrimental effect on concurrent vowel segregation when an onset asynchrony cue is also available. One explanation could be that envelope cues “distract” 3-month-old infants, counteracting the benefit of onset asynchrony cues. The possibility that an auditory cue distracts infants and might detract from their performance has been discussed in previous work (Oster & Werner, under review; Werner, 2013; Werner et al., 2009). For example, Werner et al. (2009) reported that 7-month-olds’ tone detection was not improved by the addition of an auditory cue signaling when the tone might occur. However, using the “listening window” paradigm, Werner et al. also showed that both infants and adults had come to expect that the tone would occur at a particular interval following the same cue. One explanation offered for these apparently

conflicting findings was that while the cue was clearly pointing the infants' attention to the correct point in time, any benefit of knowing when to listen was offset by the distracting effect of the cue. In the current experiment, amplitude envelopes were extracted from infant-directed speech, a stimulus that is known to attract infants' attention. The "distraction hypothesis" might be tested by varying the degree to which the auditory cue is attractive to infants.

#### IV. SUMMARY AND CONCLUSIONS

The current results provide a first account of combined onset asynchrony and envelope cue use in infants. Three-month-old infants do not show evidence of improved vowel segregation with combined cues, whereas 7-month-old infants receive a benefit similar to that seen in adults. Additionally, the contributions of the individual temporal cues to the benefit derived from the combined cues depended on the listener's age. In 3-month-olds, envelope appears to counteract the benefit of onset asynchrony, while it might be speculated that in 7-month-olds and adults the benefit of the cue combination is driven by the onset asynchrony cue.

Although providing some insights into combined temporal cue use in infants, these results generate more questions than answers. First, temporal cue use in 3-month-olds can currently not be directly compared to that in 7-month-olds and adults, given differences in the measures of sound segregation. To fully understand the development of temporal cue use in infancy, methodological and procedural changes have to be made so that more 3-month-olds are able to provide sensitivity data comparable to that obtained from older listeners. Second, the current results cannot definitively answer the question to what extent each cue contributes to the benefit of combined temporal cues. Answering this question would require the development of methods suitable to evaluate cue weighting in infants. Lastly, it is not clear how top-down mechanisms influence sound source segregation in infants. It might be assumed that adults use

different strategies than infants. For example, as soon as adults learn that the target vowel is always spoken by the male talker, adults may use that information to actively ignore the female talker. Generalizing from the fact that infants do not listen selectively in the frequency domain (Bargones & Werner, 1994), they may not use this strategy. It would be informative to alter experimental procedures to include a measure of selective auditory attention. One might, for example, include target vowels spoken by the female talker and evaluate infant and adult responses to those target vowels. If adults indeed selectively attend to the male talker and infants do not, then adults, but not infants, would be expected to respond less to target vowels spoken by the female than the male talker.

In conclusion, while the current results raise many new questions, they do suggest that immature use of temporal cues, individually or in combination, is not a major contributor to infants' difficulties segregating competing speech sounds after 7 months of age.

## CHAPTER 5 CONCLUSIONS AND FUTURE DIRECTIONS

The results of these experiments suggest that by 3-months of age, infants use onset asynchrony as a cue to sound source segregation. In fact, by 7-months of age infants appear to gain as much benefit from onset asynchrony cues as adults. In contrast, we did not find any evidence that 3-month-old infants use differences between amplitude envelopes to segregate competing sounds. Yet, by 7 months of age, infants receive as much benefit from envelope cues as adults. This difference in envelope cue use between 3- and 7-month-old infants is rather surprising, given that sensitivity to AM is similar in both age groups. This might suggest, that the ability to use envelope cues develops between 3- and 7-months of age. When onset asynchrony and envelope cues are combined, 3-month-old infants show no evidence of improved vowel segregation. Thus, the benefit 3-month-olds gained from onset asynchrony seems to have been counteracted by the presence of envelope cues. Seven-month-old infants, on the other hand, gained as much benefit from combined cues as adults. This benefit was not different from that obtained with onset asynchrony or envelope cues alone. The combined cue results are similar to those of prior research demonstrating that onset asynchrony drives the combined cue benefit (Hall et al., 2006), leading to the speculation that onset asynchrony dominated combined cue use in both 7-month-old infants and adults.

Taken together, the results from this work indicate that infants' ability to use onset asynchrony cues likely does not limit their ability to process competing sounds. In contrast, envelope cues may limit infants' concurrent sound segregation in the first 6 months of life. However, these results also generate a host of new questions that warrant further systematic investigation.

1. The procedures used here were not suitable to obtain sensitivity data from 3-month-old infants. In order to directly compare auditory cue use across development, experimental paradigms need to be developed that allow the inclusion of participants of different ages.
2. The current work provides indirect evidence that infants and adults differ in their use of voice characteristic cues. This evidence is derived from the age-related TMR difference in the absence of temporal cues. Since competing vowels were produced by a female and a male talker, it stands to reason that adults used the difference in voice characteristics to segregate the vowels even at large negative TMRs. Infants on the other hand required a large positive TMR, likely because they did not use voice characteristic cues to segregate the concurrent vowels. This conclusion is similar to that reached by two prior studies that investigated the use of voice differences in infants and toddlers (Leibold et al., 2013; Newman & Morini, 2017). However, how much voice characteristic cues contributed to infants' and adults' performance and how they interacted with the provided temporal cues, is not apparent from the current results. This question could be answered by using synthetic speech sounds and systematically varying auditory cues.
3. The results of the current work are also limited by the type of AM used. Envelopes found in naturally produced infant directed speech were chosen under the assumption that if infants use envelope cues, they would be more likely to use those that are naturally available to them. It is possible that other envelope cues would produce different results.
4. The approach used in this work did not allow for the direct comparison of individual auditory cues. How much different auditory cues are used could for example be evaluated by deriving psychometric functions or measuring weighting functions.

5. Lastly, it is not clear how auditory attention influenced the current results. For example, it might be assumed that infants and adults directed their auditory attention differently. Adults may have selectively attended to the male talker using for example fundamental frequency cues. Infants on the other hand do not listen selectively in frequency and may have distributed attention broadly to all frequency components (Bargones & Werner, 1994). One way to assess the influence of attention in this task would be to include a block of trials where the target and competing talkers switch. If infants do not direct attention in the same way as adults, they may not experience a detriment in performance, whereas adults would be expected to perform worse in the “probe attention” block.

Finally, the current work provides a basis for the evaluation of auditory cue use in infants and children with cochlear implants. While spectral and location cues are degraded in CIs, temporal cues, including onset and envelope cues, are well represented. Prior research has demonstrated that adults with CI use both temporal cues in sound segregation (Carlyon, Long, Deeks, & McKay, 2007; Chatterjee et al., 2006; Kwon & Perry, 2014; Pierzycki & Seeber, 2014). From the current results it would be expected that as long as infants and children with CI have access to onset asynchrony and envelope cues, they should be able to use them in concurrent sound segregation. Whether this is indeed the case should be evaluated in future work.

## BIBLIOGRAPHY

- Alston, E., & James-Roberts, I. (2005). Home environments of 10-month-old infants selected by the WILSTAAR screen for pre-language difficulties. *International Journal of Language & Communication Disorders*, 40(2), 123-136. doi:10.1080/13682820400006861
- Assmann, P. F., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. *The Journal of the Acoustical Society of America*, 88(2), 680-697. doi:<http://dx.doi.org/10.1121/1.399772>
- Bargones, J. Y., & Werner, L. A. (1994). Adults listen selectively; infants do not. *Psychological Science (Wiley-Blackwell)*, 5(3), 170-174.
- Bargones, J. Y., Werner, L. A., & Marean, G. C. (1995). Infant psychometric functions for detection: Mechanisms of immature sensitivity. *The Journal of the Acoustical Society of America*, 98(1), 99-111. doi:<http://dx.doi.org/10.1121/1.414446>
- Bendixen, A., Haden, G. P., Nemeth, R., Farkas, D., Torok, M., & Winkler, I. (2015). Newborn Infants Detect Cues of Concurrent Sound Segregation. *Developmental Neuroscience*, 37(2), 172-181. doi:10.1159/000370237
- Bernstein, J. G. W., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 125, 3358. doi:10.1121/1.3110132
- Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer (Version 6.0.36). <http://www.praat.org/>.
- Borrill, S. J., & Moore, B. C. J. (2002). Evidence that comodulation detection differences depend on within-channel mechanisms. *The Journal of the Acoustical Society of America*, 111(1 Pt 1), 309. doi:10.1121/1.1426373
- Bregman, A. S. (1990). *Auditory scene analysis : the perceptual organization of sound*. Cambridge, Mass. : MIT Press.
- Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, 109(3), 1101-1109. doi:10.1121/1.1345696
- Bull, D., Eilers, R. E., & Oller, D. K. (1984). Infants' discrimination of intensity variation in multisyllabic stimuli. *Journal of the Acoustical Society of America*, 76(1), 13-17. doi:10.1121/1.391110
- Buss, E., Hall, J. W., & Grose, J. H. (2012). Development of Auditory Coding as Reflected in Psychophysical Performance. In L. A. Werner, R. R. Fay, & A. N. Popper (Eds.), *Human Auditory Development* (pp. 107-136). New York: Springer.
- Cabrera, L., Bertoncini, J., & Lorenzi, C. (2013). Perception of Speech Modulation Cues by 6-Month-Old Infants. *Journal of Speech, Language & Hearing Research*, 56(6), 1733-1744. doi:10.1044/1092-4388(2013/12-0169)
- Cabrera, L., & Werner, L. (2017). Infants' and Adults' Use of Temporal Cues in Consonant Discrimination. *Ear and Hearing*, 38(4), 497-506. doi:10.1097/AUD.0000000000000422

- Carlyon, R. P., Long, C. J., Deeks, J. M., & McKay, C. M. (2007). Concurrent sound segregation in electric and acoustic hearing. *Jaro-Journal of the Association for Research in Otolaryngology*, 8(1), 119-133. doi:10.1007/s10162-006-0068-1
- Chatterjee, M., Sarampalis, A., & Oba, S. I. (2006). Auditory stream segregation with cochlear implants: A preliminary report. *Hearing Research*, 222(1-2), 100-107. doi:10.1016/j.heares.2006.09.001
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, 119(3), 1562. doi:10.1121/1.2166600
- 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:10.2307/1131286
- Demany, L. (1982). Auditory Stream Segregation in Infancy. *Infant Behavior and Development: An International and Interdisciplinary Journal*, 5(3), 261-276. doi:10.1016/S0163-6383(82)80036-2
- Eggermont, J., & Moore, J. (2012). Morphological and Functional Development of the Auditory Nervous System. In L. Werner, L. Fay, & A. Popper (Eds.), *Human Auditory Development* (pp. 61 - 105). New York, NY: Springer.
- Fogerty, D., Xu, J., & Gibbs, B. (2016). Modulation masking and glimpsing of natural and vocoded speech during single-talker modulated noise: Effect of the modulation spectrum. *J. Acoust. Soc. Am.*, 140(3), 1800-1816. doi:10.1121/1.4962494
- Folland, N. A., Butler, B. E., Payne, J. E., & Trainor, L. J. (2015). Cortical Representations Sensitive to the Number of Perceived Auditory Objects Emerge between 2 and 4 Months of Age: Electrophysiological Evidence. *Journal of cognitive neuroscience*, 27(5), 1060. doi:10.1162/jocn\_a\_00764
- Folland, N. A., Butler, B. E., Smith, N. A., & Trainor, L. J. (2012). Processing simultaneous auditory objects: infants' ability to detect mistuning in harmonic complexes. *The Journal of the Acoustical Society of America*, 131(1), 993. doi:10.1121/1.3651254
- Füllgrabe, C., Berthommier, F., & Lorenzi, C. (2006). Masking release for consonant features in temporally fluctuating background noise. *Hearing Research*, 211(1), 74-84. doi:<https://doi.org/10.1016/j.heares.2005.09.001>
- Gorga, M. P., Kaminski, J. R., Beauchaine, K. L., Jesteadt, W., & Neely, S. T. (1989). Auditory Brainstem Responses from Children Three Months to Three Years of Age: Normal Patterns of Response II. *Journal of Speech Language and Hearing Research*, 32(2), 281. doi:10.1044/jshr.3202.281
- Grimault, N., Bacon, S. P., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *The Journal of the Acoustical Society of America*, 111(3), 1340.
- Grose, J. H., & et al. (1993). Temporal Analysis in Children. *Journal of Speech and Hearing Research*, 36(2), 351-356.
- Hall, J. W., Buss, E., & Grose, J. H. (2005). Informational masking release in children and adults. *The Journal of the Acoustical Society of America*, 118, 1605. doi:10.1121/1.1992675
- Hall, J. W., Buss, E., & Grose, J. H. (2006). Comodulation detection differences for fixed-frequency and roved-frequency maskers. *The Journal of the Acoustical Society of America*, 119(2), 1021. doi:10.1121/1.2151788

- Hall, J. W., Buss, E., & Grose, J. H. (2008). Comodulation detection differences in children and adults. *The Journal of the Acoustical Society of America*, 123(4), 2213. doi:10.1121/1.2839006
- Hall, J. W., Buss, E., & Grose, J. H. (2014). Development of speech glimpsing in synchronously and asynchronously modulated noise. *The Journal of the Acoustical Society of America*, 135(6), 3594. doi:10.1121/1.4873518
- Hall, J. W., Buss, E., Grose, J. H., & Roush, P. A. (2012). Effects of age and hearing impairment on the ability to benefit from temporal and spectral modulation. *Ear and Hearing*, 33(3), 340. doi:10.1097/AUD.0b013e31823fa4c3
- Hall, J. W. I., Grose, J. H., & Dev, M. B. (1997). Auditory development in complex tasks of comodulation masking release. *Journal of Speech, Language, and Hearing Research*, 40(4), 946.
- Hedrick, M. S., & Madix, S. G. (2009). Effect of Vowel Identity and Onset Asynchrony on Concurrent Vowel Identification. *Journal of Speech, Language, and Hearing Research*, 52(3), 696-705. doi:10.1044/1092-4388(2008/07-0094)
- Hill, N., & Darwin, C. J. (1996). Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning. *Journal of the Acoustical Society of America*, 100(4), 2352-2364.
- Horn, D., Hutton Gerhards, C., & Werner, L. (2016). *Temporal Modulation Cutoff Frequency in Normal Hearing Infants*. Paper presented at the Association for Research in Otolaryngology, San Diego.
- Kuhl, P. K. (1979). Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America*, 66(6), 1668. doi:10.1121/1.383639
- Kwon, B. J., & Perry, T. T. (2014). Identification and Multiplicity of Double Vowels in Cochlear Implant Users. *Journal of Speech, Language, and Hearing Research*, 57(5), 1983-1996. doi:10.1044/2014\_JSLHR-H-12-0410
- Lau, B. K., Lalonde, K., Oster, M.-M., & Werner, L. A. (2017). Infant pitch perception: Missing fundamental melody discrimination. *The Journal of the Acoustical Society of America*, 141(1), 65-72. doi:10.1121/1.4973412
- Leibold, L. J. (2012). Development of Auditory Scene Analysis and Auditory Attention. In L. A. Werner, Fay R. R., Popper, A. N. (Ed.), *Human Auditory Development* (pp. 137-161). New York: Springer.
- Leibold, L. J. (2017). Speech Perception in Complex Acoustic Environments: Developmental Effects. *Journal of Speech, Language, and Hearing Research*, 60(10), 3001-3008. doi:10.1044/2017\_JSLHR-H-17-0070
- Leibold, L. J., Bonino, Y. A., & Buss, E. (2016). Masked Speech Perception Thresholds in Infants, Children, and Adults. *Ear and Hearing*, 37(3), 345. doi:10.1097/AUD.0000000000000270
- Leibold, L. J., & Neff, D. L. (2007). Effects of masker-spectral variability and masker fringes in children and adults. *Journal of the Acoustical Society of America*, 121(6), 3666-3676. doi:10.1121/1.2723664
- Leibold, L. J., Taylor, C., Hillock-Dunn, A., & Buss, E. (2013). Effect of talker sex on infants' detection of spondee words in a two-talker or a speech-shaped noise masker. *J. Acoust. Soc. Am. The Journal of the Acoustical Society of America*, 133(5), 3338.

- Leibold, L. J., & Werner, L. A. (2006). Effect of masker-frequency variability on the detection performance of infants and adults. *Journal of the Acoustical Society of America*, *119*(6), 3960-3970. doi:10.1121/1.2200150
- Lentz, J. J., & Marsh, S. L. (2006). The effect of hearing loss on identification of asynchronous double vowels. *Journal of speech, language, and hearing research : JSLHR*, *49*(6), 1354.
- Lewkowicz, D. J., & García Coll, C. (2010). Infant Perception of Audio-Visual Speech Synchrony. *Developmental Psychology*, *46*(1), 66-77. doi:10.1037/a0015579
- Mayo, C., & Turk, A. (2005). The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception. *The Journal of the Acoustical Society of America*, *118*(3), 1730-1741. doi:doi:<http://dx.doi.org/10.1121/1.1979451>
- McAllister, A. M., Granqvist, S., Sjölander, P., & Sundberg, J. (2009). Child voice and noise: a pilot study of noise in day cares and the effects on 10 children's voice quality according to perceptual evaluation. *Journal of voice : official journal of the Voice Foundation*, *23*(5), 587. doi:10.1016/j.jvoice.2007.10.017
- Miller, G. A., & Licklider, J. C. R. (1950). The Intelligibility of Interrupted Speech. *Journal of the Acoustical Society of America*, *22*(2), 167-173. doi:10.1121/1.1906584
- Moore, B. C. J., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B*, *367*(1591), 919-931. doi:10.1098/rstb.2011.0355
- Moore, J. K., & Linthicum, F. H. (2007). The human auditory system: a timeline of development. *International Journal of Audiology*, *46*(9), 460.
- Newman, R. S. (2005). The cocktail party effect in infants revisited: Listening to one's name in noise. *Developmental Psychology*, *41*(2), 352-362. doi:10.1037/0012-1649.41.2.352
- Newman, R. S., & Morini, G. (2017). Effect of the relationship between target and masker sex on infants' recognition of speech. *Journal of the Acoustical Society of America*, *141*(2), EL164-EL169. doi:10.1121/1.4976498
- Nie, Y., & Nelson, P. (2015). Auditory stream segregation using amplitude modulated bandpass noise. *Frontiers in Psychology*, *6*(1151). doi:10.3389/fpsyg.2015.01151
- Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *Journal of the Acoustical Society of America*, *115*(4), 1777-1790. doi:10.1121/1.1651192
- Nittrouer, S., & Lowenstein, J. H. (2014). Dynamic Spectral Structure Specifies Vowels for Adults and Children. *Language and Speech*, *57*(4), 487-512. doi:10.1177/0023830913508075
- Nittrouer, S., Lowenstein, J. H., & Tarr, E. (2013). Amplitude Rise Time Does Not Cue the /ba/-/wa/ Contrast for Adults or Children. *Journal of Speech Language and Hearing Research*, *56*(2), 427-440. doi:10.1044/1092-4388(2012/12-0075)
- Nozza, R. J. (1988). Binaural Release from Masking for a Speech Sound in Infants, Preschoolers, and Adults. *Journal of Speech and Hearing Research*, *31*(2), 212-218.
- Nozza, R. J., Rossman, R. N. F., Bond, L. C., & Miller, S. L. (1990). Infant speech-sound discrimination in noise. *Journal of the Acoustical Society of America*, *87*(1), 339-350.
- Nozza, R. J., & Wilson, W. R. (1984). Masked and unmasked pure-tone thresholds of infants and adults - development of auditory frequency-selectivity and sensitivity *Journal of Speech and Hearing Research*, *27*(4), 613-622.
- Oster, M.-M., & Werner, L. A. (2017). The Influence of Target and Masker Characteristics on Infants' and Adults' Detection of Speech. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3625-3631. doi:10.1044/2017\_JSLHR-H-16-0464

- Pierzycki, R. H., & Seeber, B. U. (2014). Comodulation Masking Release In Electric Hearing. *Journal of the Association for Research in Otolaryngology*, *15*(2), 279-291. doi:10.1007/s10162-013-0433-9
- Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221-231. doi:[http://dx.doi.org/10.1016/S0167-6393\(02\)00105-X](http://dx.doi.org/10.1016/S0167-6393(02)00105-X)
- Ponton, C. W., Moore, J. K., & Eggermont, J. J. (1996). Auditory brain stem response generation by parallel pathways: differential maturation of axonal conduction time and synaptic transmission. *Ear and Hearing*, *17*(5), 402. doi:10.1097/00003446-199610000-00006
- Ramirez-Esparza, N., Garcia-Sierra, A., & Kuhl, P. K. (2017). Look Who's Talking NOW! Parentese Speech, Social Context, and Language Development Across Time. *Front. Psychol.*, *8*. doi:10.3389/fpsyg.2017.01008
- Seidl, A., & Cristia, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental science*, *11*(4), 596-606. doi:10.1111/j.1467-7687.2008.00704.x
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in cognitive sciences*, *12*(5), 182-186. doi:10.1016/j.tics.2008.02.003
- Singh, P. G., & Bregman, A. S. (1997). The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *The Journal of the Acoustical Society of America*, *102*(4), 1943.
- Sinnott, J. M., & Aslin, R. N. (1985). Frequency and intensity discrimination in human infants and adults. *Journal of the Acoustical Society of America*, *78*(6), 1986-1992. doi:10.1121/1.392655
- Stuart, A. (2005). Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise. *Ear and Hearing*, *26*(1), 78. doi:10.1097/00003446-200502000-00007
- Stuart, A. (2008). Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children. *Journal of the American Academy of Audiology*, *19*(2), 135-146.
- Summerfield, Q., Culling, J. F., & Fourcin, A. J. (1992). Auditory Segregation of Competing Voices: Absence of Effects of FM or AM Coherence [and Discussion]. *Philosophical Transactions: Biological Sciences*, *336*(1278), 357-366.
- Sussman, E., Wong, R., Horvath, J., Winkler, I., & Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5-11-year-old children. *Hearing Research*, *225*(1-2), 117-127. doi:10.1016/j.heares.2006.12.013
- Valentine, S., & Lentz, J. J. (2012). The Influence of Reduced Audible Bandwidth on Asynchronous Double-Vowel Identification. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1750-1764. doi:10.1044/1092-4388(2012/11-0033)
- Werner. (2017). Infants and children at the cocktail party. In J. Middlebrooks, Simon, J. and Popper, AN (Ed.), *The auditory system at the cocktail party*: Springer.
- Werner, L. (2006). Amplitude modulation detection by infants and adults. *The Journal of the Acoustical Society of America*, *119*(5), 3234-3234. doi:10.1121/1.4785975
- Werner, L. A. (1995). Observer-based approaches to human infant psychoacoustics. In G. M. Klump, R. J. Dooling, R. R. Fay, & W. C. Stebbins (Eds.), *Biomethods; Methods in comparative psychoacoustics* (Vol. 6, pp. 135-146).
- Werner, L. A. (2013). Infants' detection and discrimination of sounds in modulated maskers. *Journal of the Acoustical Society of America*, *133*(6), 4156-4167. doi:10.1121/1.4803903

- Werner, L. A., & Bargones, J. Y. (1991). Sources of auditory masking in infants: distraction effects. *Perception & Psychophysics*, 50(5), 405-412.
- Werner, L. A., & Boike, K. (2001). Infants' sensitivity to broadband noise. *Journal of the Acoustical Society of America*, 109(5), 2103-2111. doi:10.1121/1.1365112
- Werner, L. A., Folsom, R. C., Mancl, L. R., & Syapin, C. L. (2001). Human Auditory Brainstem Response to Temporal Gaps in Noise. *Journal of Speech, Language, and Hearing Research*, 44(4), 737.
- Werner, L. A., Parrish, H. K., & Holmer, N. M. (2009). Effects of temporal uncertainty and temporal expectancy on infants' auditory sensitivity. *Journal of the Acoustical Society of America*, 125(2), 1040-1049. doi:10.1121/1.3050254
- Winkler, I., Kushnerenko, E., Horváth, J., Čeponienė, R., Fellman, V., Huotilainen, M., Näätänen, R., & Sussman, E. (2003). Newborn Infants Can Organize the Auditory World. *Proceedings of the National Academy of Sciences of the United States of America*, 100(20), 11812-11815.