Knowing Your Name: Hearing and Recognition in Infants

Candace S. Ali
Mentor: Dr. Rochelle Newman, Professor of Hearing and Speech Sciences
University of Maryland, College Park

Abstract

Infants are constantly placed in situations where they are exposed to multiple sources of sound including orators, music, television, or other causes of background noise. Young infants have the ability to separate streams of speech and selectively attend to speech signals, however little is known about the performance of infants and the cues used to recognize speech in the presence of background noise. In order to comprehend language development in infants, we must be able to understand how they acquire language despite noisy situations. This quantitative study seeks to examine how well infants from 3.5 – 5.5 months pay attention and process speech when in the presence aof competing background noise. The head-turn preference procedure was conducted, and stimuli were created based on the infants own name and a foil name, presented in constant and varying-amplitude noise in order to test the infants' ability to recognize their own name (a familiar word) despite the opposing noise. Preliminary results indicate that infants displayed a preference for their own name under the constant-amplitude condition only . This suggests that the varying-amplitude condition is too distracting for infants to recognize speech and may not be as useful for their language acquisition in comparison to constant-amplitude. However, due to the lack of participation, further testing must continue to gain significant results and to draw further conclusions among infants of this age group.

Introduction

Problem Statement

How well do infants pay attention to a speaker when other acoustic elements are present? Research shows that young infants, 1 year or less, are limited in their speech recognition unless the background noise was 10 dB less intense than the speaker (Newman, 2005). If signal-to-noise- ratios are closer together, infants will have difficulty attending to speech under noisy conditions. For young infants to recognize speech in environments where there are multiple talkers, background music, or television sets, the signal-to-noise ratio must be significantly higher than for adult listeners (Newman, 2009). Yet, this may not be accomplished in real-world situations. Therefore, there is a need to examine how well infants recognize and process speech when in the presence of background noise using settings which are more typical in their everyday life.

Purpose of Study

In order to fully understand language development in infants we must understand their speech recognition ability in the presence of noise. Studies have been conducted in quiet environments; however, infants frequently find themselves in the presence of noise and are then given greater opportunities to separate speech signals. Infants that are able to separate voices and attend selectively to a particular speaker could be at an advantage for learning language than infants who are unable to do so (Newman, 2005). Understanding speech signals in noise could strengthen infant's ability to acquire language. They are able to group and identify differences in acoustic properties, while choosing to pay attention to the target signal over the distracter signal. As we identify streaming in infants, we can assume that those able to segregate speech signals are using cognitive processes and those unable to succeed in these areas may have difficulty identifying speech in the presence of noise.

The current study will identify whether infants 4 months of age are able to recognize speech in the presence of competing amplitude-modulated background noise. Using a common word, their own name, we will identify whether or not they show a preference for their own name over the foil name in the presence of different noise environments. In order to learn language in noisy situations infants must have the capacity to stream speech signals, and process the necessary information to gain an understanding of what is being stated. The goal of this research is to understand what acoustic environments are best for infants recognize speech signals.

Overall, if infants are able to listen longer to their name in noisy settings we can confirm that those situations are not too distracting for them to segregate speech signals. Based on the length of time that each infant spent looking towards the sound source during the head- turn preference procedure we can identify that they were able to pay attention despite the constant or varying noise backgrounds.

Research Questions

In this study there are several research questions we will address. First, can an infant differentiate his/her own name versus a foil name with competing background noise? Second, what type of noise is "easier" for infants to stream speech signals: constant amplitude or varying amplitude? Lastly, will the infant be distracted by the background noise? In this study, we will examine how well 4 month old infants listen and recognize their own name with amplitude-modulated noise at 5 Hz, as well as how infants pay attention and learn from speech under noisy conditions.

Delimitations of Research

English-only speakers were not the sole participants in this study because the child's own name was used. Therefore, language was not a factor for eligibility of the study because the child could respond to their own name regardless of language barriers. Infants of 3.5 – 5.5 months were used to better understand the early stages of language development and speech recognition. Children with hearing impairments were not used because we wanted to study normal hearing infants and identify their reactions to the different noise conditions.

Definitions Related to Research

There are several terms that will be discussed, including: streaming, selective attention, head turn preference procedure (HPP), foil name, response box, infant directed speech (IDS), signal-to-noise ratio, amplitude modulated, and masking.

Streaming. The ability to group together sounds and separates them from sounds that are coming from another source (Gleason & Ratner, 1998).

Selective Attention. The ability to choose which items to stream and what sounds you will pay attention to.

Head-turn Preference Procedure (HPP). This procedure is conducted within a three sided pegboard booth. "Loudspeakers are mounted into the walls of the two side panels at about the level of the infant's head. A small red light is mounted on each of the side panels in the vicinity of the loudspeaker. The center panel which the infant faces has a small green light mounted at the infant's eye level" (Nelson, Jusczyk, Mandel, & Myers, 1995, p.112)." Trials begin by drawing infants attention to the center light. Once the infant attends to the light, it is turned off and a flashing light appears on one of the side panels. "Once the infant turns to that side, the stimulus begins to play. It continues (and the side light keeps flashing) until the infants turns away for at least a continuous period of 2 s (or until the entire stimulus for that trial has been played. The infants looking time is the total time the infant orients to the sample (Nelson et. al., 1995, p.112)."

Foil Name. An unfamiliar name is created and matched for syllable stress based on infant's real name or name most commonly called.

Response Box. A button box attached to the computer in the experimenter booth, which is used to input data and to code the infants behavior during the experiment. The response box includes right, left, away, and center keys to code for the infants head turns during the study.

Infant Directed Speech (IDS). Refers to the intonation, stress, pitch, and overall manner that we use to speak to infants rather than adults (Gleason & Ratner, 1998).

Signal-to-Noise Ratio. The ratio of the signal to the corresponding noise (Yost, 2007).

Amplitude-Modulation. The encoding of a carrier wave by variation of its amplitude in accordance with an input signal (Gelfand, 2009). For the present study, there was not an input signal, and the carrier wave varied over time, sinusoidally.

Masking. The process by which the threshold of audibility for one sound is raised by the presence of another (masking) sound (Yost, 2007).

Review of the Literature

The following section reviews empirical studies that address speech recognition, auditory stream segregation, and the signal-to-noise ratio. The review of literature supports the need for additional research to add to the body of knowledge of speech segregation in infants and to understand the environments necessary to support positive language acquisition.

Speech Recognition in Noise

Developing infants are learning to selectively attend to speech; however they may find it difficult to do so in settings that have more than one auditory signal. Newman (2005) suggests that infants at 5 months are unable to attend to their own names when the intensity of the background noise increases and is relatively close to the level of the speaker. Nevertheless, infants at this age show a capacity for speech discrimination, but they are still limited in their ability to separate auditory signals.

Infants are better equipped to segregate speech signals when there are multiple-talkers rather than single-talkers in the background; however, this is the opposite separation pattern of adults (Newman, 2009). Single-talker settings vary in amplitude level over time, and are easier for adults to recognize speech. A single voice has much more varying-amplitude than does a combination of voices. When there is multi-talker babble, these variations average out. Newman (2009) identified that time-varying acoustic properties in single voice conditions are more distracting to infants than multiple voice conditions and infants were able to hear their own name under competing multi-talker conditions rather than single-talker. Interestingly, the authors state that this finding could be due to the acoustic properties of the single-voice background, either attracting the infants' attention or serving as a masker for the speech stream (Newman, 2009).

Auditory Stream Segregation

Auditory stream segregation, known as streaming is defined as the situation in which a "temporal sequence of sounds is organized into two or more perceptual auditory entities or 'streams,' which can be individually followed by listeners and are generally considered arising from different acoustic sources (Miller and Heise, 1950; Bregman, 1978; as cited in Grimault, Bacon & Micheyl, 2002, p.1340)." In other words, listeners are able to separate sound signals that are coming from two separate sound sources. Adult listeners find it easier to separate multiple streams of speech if they are from separate locations in space, speakers have different fundamental frequencies or genders, or visual facial information is presented from the speaker that differentiate the voices (Broadbent, 1954; Cherry, 1953; Poulton, 1953; Brokx & Nooteboom, 1982; Darwin & Hukin, 2000 as cited in Barker & Newman, 2004). However, stream segregation is a task that can be quite difficult for adults if these cues are not present (Plyler, Bahng, & Von Hapsburg, 2008). Likewise, these cues may be beneficial for both adults and infants. Yet, little information has been gathered on what ages infants are able to use these cues and limited research identifies the necessary cues that enable infants to stream speech signals.

Signal-To-Noise-Ratio

Infants require a "greater signal-to-noise-ratio than adults to achieve a given level of performance" (Nozza, Rossman, Bond, & Miller, 1990, p.340). If the background noise and speaker are too close in frequency and noise level, the infant will not be able to identify differences among the two acoustic signals, and will not show a preference for one or the other. Newman and Jusczyk (1996) identified that infants can selectively attend to a specific talker when that talker is more intense than the background speaker or when the speakers are of different genders or differ in other ways (i.e. tone, pitch). Overall, infants are more likely to have difficulty attending to speech signals if the background noise level is closer to the level of the orator.

Summary and Implications of the Literature

Recognition and speech perception have been studied in infants for many years. The information gathered identifies that there are still gaps in the research and there is not a clear understanding of when and how infants recognize speech in noisy environments and what cues may help or hinder their speech discrimination. The developmental timeline of infants' ability to listen and detect speech in the presence of noise is still in the works, and further research is needed to understand what settings are best to assist in the development of language. Infants need to be able to separate speech from background noise in order to learn language, especially considering that their common environments involve the presence several acoustic sources. Infants and adults differ in their abilities to recognize speech in the presence of noise, and there are several possibilities for these differences. We hope to uncover what is causing this difference in recognition patterns. Infants are clearly still developing and their auditory and language skills are not as matured as adults, but we would like to identify what stages in their development that they are able to accomplish speech recognition and what components assist in their development of language acquisition.

Research Design and Methodology

Participants

Infants were recruited from brochures sent out to parents who had newborns and those that responded were placed in the laboratory database. Research assistants then called and scheduled appointments for infants that were allocated to Dr. Rochelle Newman's laboratory and were the appropriate age for the study. A total of nine infants participated in this study as of June 23, 2010 (3 females, 6 males). The children were all between 3.5 – 5.5 months of age. Upon arrival, caregivers were given two consent forms (one for keep for their records and one for the research laboratory) indicating the child was allowed to participate as well as be videotaped during the session. Caregivers were asked to fill out a questionnaire regarding the infant's language and developmental history at the time of the study. Also, biographical information was requested in order to ensure diversity among participants. This included race, ethnicity, and educational background of each parent. Due to fussiness or crying, some children were unable to complete the study and were not included in the data set.

Stimuli

A target speech stream and a distracter speech stream were recorded prior to the infants scheduled appointment. Recordings were made in a sound attenuated room, and were amplified, recorded at 44.1 kHz sampling rate(digitized by a 16-bit analog-to-digital converter), and then stored on a computer disk. Foil names were created based on the syllable stress of the child's name, or commonly known name. Only female voices were used in the recordings, and only one speaker recorded for each participant. Infant-directed speech was used to record stimuli. Two practice stimuli (musical passages) were created to orient participants to the task.

Apparatus

The experiment was conducted within a three-sided booth consisting of white pegboard panels (4ft x 6ft). Side panels consisted of a red light and loudspeaker located behind the pegboard. The center panel had also a light and small hole for the video camera, which recorded all the sessions. Behind the center panel the experimenter observed the session through a monitor and coded the infant's behavior on the computer, a Macintosh G3. Coding identifies whether or not the child was looking at the sound source or not. Coding was conducted through a response box during the sessions to signal the initiation and end of flashing lights on either right, left, or center panels. The computer controlled the presentation of stimuli based on the coding of the experimenter. Curtains were hung from the ceiling to top of the test booth and back side of the booth was closed by a black curtain to eliminate all other distractions of the room.

Figure 1. Testing booth, where caregiver and infant would sit during the study.

Figure 2. Experimenter booth, experimenter sits and codes the behavior of the infant on the computer through a response box. The tester used the television set above to observe the infant.

Procedure

The caregiver was instructed to hold the infant on their lap while sitting in the center of the test booth. The head-turn preference procedure was used during the study (Nelson et. al., 1995). To reiterate, every trial began with a blinking light in the center of the front panel in order to attract the infant's attention. Once the infant was acquainted with this signal, the flashing light was shut off and one of the side lights of either two panels began to flash. If the infant oriented to the flashing light of either side panel then the loudspeaker on that side would also begin playing the stimuli. The stimulus would continue playing until its completion or until the child looked away for at least 2 consecutive seconds. Listening time was based on the amount of time the infant spent looking at the flashing light. Data collection included the duration of the stimuli, the infants looking time at the flashing light, and the direction of head turns. All of these specific details were recorded on the computer during the experiment.

There were two sections of this experiment. A practice phase and a test phase. The practice phase familiarized the infants with the task before the actual test began. The infants heard musical passages, which were 14.9 seconds long. Infants had to acquire at least 25 seconds of listening time to each musical selection in order to move into the test phase. Once the listening criterion was achieved the test phase began. The test phase included the foil name and the infants' name in constant and varying noise. There were 16 trials separated into 4 blocks. Each block consisted of one of the four conditions: foil name in constant noise, foil name in varying amplitude noise, own name in constant noise, and own name in varying noise.

The experimenter used a response box to indicate whether the infant looked center, right, left, or away from the source (flashing light). The experimenter and caregiver listened to masking music over headphones during the experiment in order to prevent bias during the test or during coding responses.

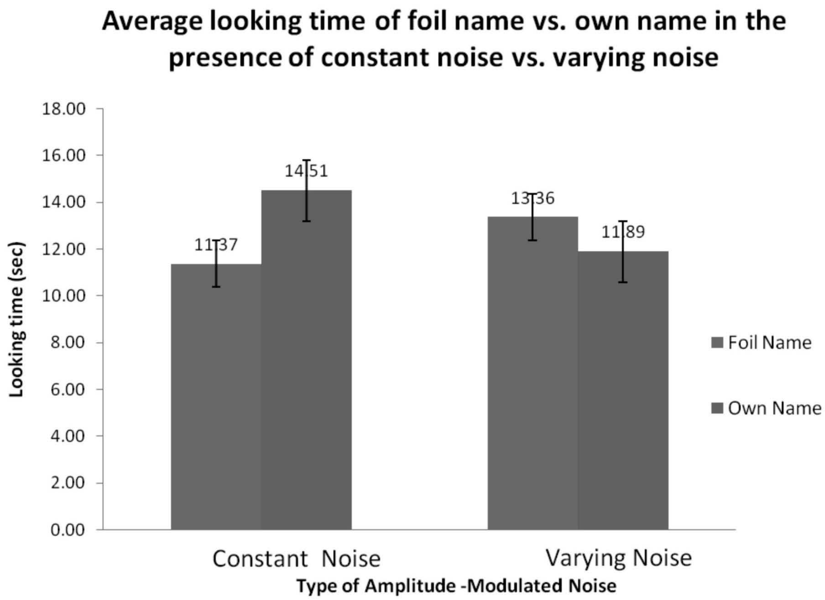
Data Collection

Foil/ Name Varying Noise vs. Foil/ Name Constant Noise				
Identifier	Foil Constant	Name Constant	Foil Varying	Name Varying
2IJ	6.85	13.86	16.10	10.80
3CR	12.99	8.37	11.38	11.66
4OB	8.52	8.05	3.92	5.61
5AP	12.52	10.95	17.03	10.52
6JL	4.77	9.16	3.72	8.28
8LD	7.73	10.02	10.02	8.92
9MS	16.43	25.30	25.30	17.22
11TM	17.92	22.95	19.56	10.35
13WH	14.56	21.90	13.24	23.67
N= 9				
Mean	11.37	14.51	13.36	11.89
Standard Deviation	4.58	6.92	7.06	5.41
Comparison of name and foil in constant noise				
Probability	0.07			
T-Value	2.06			
Comparison of name and foil in amplitude-varying noise				
Probability	0.51			
T-Value	0.68			

Table 1. Raw data of looking times for each participant of the study from May 2010 to June 2010.

Note: Identifier refers to participant number, based on the order that each participant was scheduled for the study along with the initials of their first and last name.

Figure 3. Mean looking time of 9 participants for own name and foil name conditions in the presence of constant and varying noise. Looking times ranged from 11.37sec to 14.51sec, with the constant own name condition displaying highest average.



Data Analysis Strategies

Currently, data was analyzed using t-tests comparing the name constant to the foil constant and name varying to the foil varying conditions in order to identify whether infants listened more to their own name, than the foil name in each type of noise. This technique was used to relate the size of the difference to the amount of variation among the infants and determine a consistent pattern. Figure 3 was created to observe the average looking times as well as overall preferences (if any) among the four conditions.

Strategies for Minimizing Bias and Error

To prevent the speaker from recording the infants' own name and foil name in a manner that would be more appealing to the infant, speakers were not told which were target or foil names. During the testing session, experimenters were aware of which names were the targets. However, both the experimenter and caregiver wore Peltor aviation headphones and listened to masking music during the session to avoid influencing the child's behavior or coding of head movements.

Results, Conclusions, and Recommendations for Future Research

Results

In figure 3, the mean listening times were calculated for the four different conditions. Infants averaged 14.51 seconds of listening to their own names in the constant noise while the foil name only averaged 11.37 seconds of listening time. However, in the varying noise condition, the own name stimuli averaged at 11.89 seconds while the foil name averaged at 13.36 seconds. Though there is not a large enough sample for significant findings, we can identify that participants are recognizing their own name more so in the presence of constant noise than in varying noise. We found that infants listened longer to their own name in the presence of constant noise [$t(8) = 2.06$, $p = 0.07$] but failed to so under the when the background consisted of a varying amplitude [$t(8) = 0.68$, $p = 0.51$].

In table 1, the raw data is presented to identify the exact listening times at which the participants responded. The data shows no significant results. Although, there is a small difference between the constant noise own name condition and the varying noise foil name condition, the results display a preference for the own name condition in constant noise.

Limitations of the Study

In the present study infants were not provided visual cues. Infants were unable to make use of visual cues by seeing the expression of the speakers' face, which could aid in the separation of speech and noise (Hollich, Newman, & Jusczyk (in press)). However, because the use of visual information was not provided for participants we cannot identify whether or not infants would respond to their name in the presence of varying noise. We did not compare whether children are normally placed in quiet or noisy living conditions. Therefore, the study cannot identify if those infants who are more likely surrounded by noisy environments on a daily basis were able to stream better in comparison to children who are more exposed to quiet settings. There could be differences among children that are raised in noisy environments versus children raised in quiet environments, however these differences were not acknowledged in this study. Measures of the daily signal-to-noise ratios for each infant were not identified and could not be used to explain the results of the current study. Thorough data was not gathered on the progression of the infants' language and overall maturation, thus developmental factors were not considered in addressing the results of the study. Cultural accents were not identified during the time of the study. Participants who are spoken to under living conditions with accents from other cultures may respond differently when hearing there name with infant directed speech and an American accent. However, we did not identify these differences among participants; therefore cross-cultural differences were not gathered. Also, because additional age groups were not used in this study, comparisons among different ages could not be made at the time of the experiment.

Conclusions

The results of the present experiment suggest that infants are able to identify speech in the presence of constant noise rather than varying noise. The purpose of this study was to identify how well developed speech recognition was in infants in the presence of competing background noise consisting of two different noise conditions. Drawing on the findings at this point during the study we can detect that infants are showing a preference for their own name in the presence of constant noise, therefore the constant noise was not too distracting for them to recognize speech signals. Infants cannot utilize varying noise and instead must try harder to perceive the speech signals. Overall, infants were able to push past the noise and recognize speech better in constant noise. However, because the sample set consisted of nine participants, the data collected shows no significant findings. In order to expand on this topic, the sample size will need to increase in order to have substantial data set for further conclusions.

Recommendations for Future Research

Certain participant names may be more complex and have more syllables than other participants. Simple names with fewer syllables may be easier for infants to recognize in the presence of noise. The simplicity of names may be the cause of the differences among the looking times of some participants in certain conditions rather than others. In future studies, perhaps more complex names can be compared to infants with less difficult names in the presence of noise. Also, the personal exposure levels of quiet and noisy environments were unknown. Perhaps by identifying the environments that infants' are more or less exposed to on a daily basis we can then get a better understanding for the preliminary results found. Infants exposed to noisier conditions may have the advantage over infants that are not exposed to those same situations. In addition, future studies are needed to address the changes over the developmental stages of infants and identify when they are able to understand speech signals under different noise conditions.

References

Albin, D., & Echols, C. (1996). Stressed and word-final syllables in infant-directed speech. *Infant Behavior & Development*, *19*(4), 401-418.

Bargones, J., & Werner, L. (1994). Adults listen selectively; infant do not. *Psychological Science*, *5*(3), 170-174.

Barker, B., & Newman, R. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*(2), B45-B53.

Berko Gleason, J. (Ed.), & Ratner, N. (1998). *Psycholinguistics* (2nd edition). Thompson Learning.

Cooper, R., & Aslin, R. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584-1595.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, *27*(2), 209-221.

Grimault, N., Bacon, S., & Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *The Journal Of The Acoustical Society Of America*, *111*(3), 1340-1348.

Gelfand, S. (2009). *Essential of audiology* (3rd edition). Thieme Medical Publishers.

Hollich, G.J., Newman, R.S., Jusczyk, P.W. (in press). Infants' use of synchronized visual information to separate streams of speech. *Child Development*.

Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, *3*(9), 323-328.

Kirchhoff, K., & Schimmel, S. (2005). Statistical properties of infant-directed versus adult-directed speech: insights from speech recognition. *The Journal Of The Acoustical Society Of America*, *117*(4 Pt 1), 2238-2246.

Lecanuet, J., & Granier-Deferre, C. (1993). Speech stimuli in the fetal environment. *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 237-248). New York, NY US: Kluwer Academic/Plenum Publishers.

McArdle, R., Wilson, R., & Burks, C. (2005). Speech Recognition in Multitalker Babble Using Digits, Words, and Sentences. *Journal of the American Academy of Audiology*, *16*(9), 726-739.

Merriman, W., & Stevenson, C. (1997). Restricting a familiar name in response to learning a new one: Evidence for the mutual. *Child Development*, *68*(2), 211.

Nelson, D., Jusczyk, P., Mandel, D., & Myers, J. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior & Development*, *18*(1), 111-116.

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

Newman, R.S. (2005). The cocktail party effect in infants revisited: Listening to one's name in noise. *Developmental Psychology*, 41, 352-362.

Newman, R.S. (2009). Infants' listening in multitalker environments: Effect of the number of background talkers. *Attention, Perception, & Psychophysics*. 71, 822-836.

Newman, R., Ratner, N., Jusczyk, A., Jusczyk, P., & Dow, K. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*, *42*(4), 643-655.

Nozza, R., Rossman, R., Bond, L., & Miller, S. (1990). Infant speech-sound discrimination in noise. *The Journal Of The Acoustical Society Of America*, *87*(1), 339-350.

Pittman, A., & Wiley, T. (2001). Recognition of Speech Produced in Noise. *Journal of Speech, Language & Hearing Research*, *44*(3), 487.

Plyler, P., Bahng, J., & Von Hapsburg, D. (2008). The Acceptance of Background Noise in Adult Cochlear Implant Users. *Journal of Speech, Language & Hearing Research*, *51*(2), 502-515.

Singh, L., Morgan, J., & White, K. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory & Language*, *51*(2), 173-189.

Singh, L., Nestor, S., & Bortfeld, H. (2008). Overcoming the Effects of Variation in Infant Speech Segmentation: Influences of Word Familiarity. *Infancy*, *13*(1), 57-74.

Slaney, M., & McRoberts, G. (2003). BabyEars: A recognition system for affective vocalizations. *Speech Communication*, *39*(3-4), 367-384.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501-532.

Weppelman, T., Bostow, A., Schiffer, R., Elbert-Perez, E., & Newman, R. (2003). Children's use of the prosodic characteristics of infant-directed speech. *Language & Communication*, *23*(1), 63-80.

Yost, W. (2007). *Fundamentals of hearing: an introduction* (5th edition). Academic Press.