Proceedings of Meetings on Acoustics

Volume 19, 2013

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 2aSC: Linking Perception and Production (Poster Session)

2aSC30. Effect of talker sex on infants' detection of spondee words in a two-talker or a speech-shaped noise masker

Lori J. Leibold*, Crystal N. Taylor, Andrea Hillock-Dunn and Emily Buss

*Corresponding author's address: University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, leibold@med.unc.edu

Speech recognition performance in the presence of competing speech is typically better for adults when the target and masker talkers are different sexes than when the target and masker talkers are the same sex. One explanation for this result is that the acoustic differences between male and female speech productions promote segregation of the two streams of speech, thus leading to a reduction in informational masking. In this study, an observer-based psychophysical procedure was used to compare infants' (7-13 months) masked speech detection thresholds for spondee words produced by a male or a female talker in either a two-female-talker or a speech-shaped noise masker. Infants were assigned to a single testing condition. Maskers were presented continuously throughout testing at an overall level of 50 dB SPL, fixed throughout testing. Following training to an 80%-correct criterion, thresholds for the target word were measured adaptively using a 2-down, 1-up procedure. Preliminary data indicate that infants' thresholds were similar for the female and male target words in both the two-female-talker and the speech-shaped noise masker.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Speech recognition performance in the presence of competing speech is typically better for adults when the target and masker talkers are opposite sexes than when the target and masker talkers are the same sex (e.g., Festen & Plomp, 1990; Brungart, 2001; Brungart & Simpson, 2002). One explanation for this result is that the acoustic differences between the male and female speech productions promote segregation of the two streams of speech, thus leading to a reduction in informational masking (e.g., Freyman et al. 2004).

This study examined whether infants, like adults, benefit from the mismatch in talker sex in the context of a speech detection task. The developing infant must learn about speech and language in environments containing multiple sources of competing sounds. Mounting evidence indicates that these environments offer a greater challenge for infants than for adults (e.g., Trehub et al., 1981; Newman & Jusczyk, 1996; Newman, 2009). The sources of infants’ increased susceptibility to masking are not fully understood, but do not appear to be the result of an immature peripheral auditory system, at least for infants older than about six months of age (reviewed by Buss et al., 2012). Instead, maturation of central auditory processes such as the ability to segregate and attend to targets sounds embedded in background maskers are believed to play a major role in the development of masked speech perception.

It is not known whether or not infants will benefit from the presentation of opposite-sex compared to same-sex talkers. Infants are more susceptible to informational masking compared to adults (e.g., Leibold & Werner, 2006). If the sex-mismatch cue aids in the segregation of the target words from the competing masker for infants, the potential benefit may be substantial. Alternatively, infants may demonstrate a different pattern of results relative to adults if their ability to perform sound source segregation is immature. Evidence supporting this prediction was reported by Newman (2009), who found that two age groups of infants (5.5 and 8 months) recognized their name presented at a +10 dB signal-to-noise ratio in the presence of multi-talker babble, but failed to recognize their own name presented at the same signal-to-noise ratio in either single-talker speech or time-reversed, single-talker speech. The reverse pattern of results has been reported for adults (e.g., Bronkhorst & Plomp, 1992). Newman (2009) suggested that these findings may reflect infants’ immature ability to segregate the target speech from the acoustically-similar masker speech.

METHODS

Listeners

Listeners were 7- to 13-month-old infants and adults. Selection criteria for infant listeners were: (1) no risk factors for hearing loss as assessed by parental report; (2) English spoken in the home; (3) no more than two episodes of otitis media; (4) not under treatment for otitis media within the prior week; and (5) healthy on the test date. In addition, screening tympanometry was performed on every infant participant at the end of each testing session. Peak admittance of at least 0.2 mmhos at a pressure between -200 and 50 daPa was required to pass the screening. Selection criteria for adult listeners were: (1) no risk factors for hearing loss as assessed by self-report; (2) no more than 2 years of musical training; (3) no previous participation in psychoacoustic studies; and (4) native speakers of English. In addition, adults were required to pass a hearing screening prior to testing, with thresholds less than or equal to 20 dB HL for octave frequencies between 250 and 8000 Hz (ANSI, 2010).

Stimuli

The target stimuli were the spondee words ‘hotdog’ and ‘playground’. Both spondee words were recorded in isolation from one adult female and one adult male speaker using a condenser microphone (AKG212C1000S) mounted approximately six inches from the speaker’s mouth. Productions were amplified (TDT MA3) and digitized at a resolution of 32 bits and a sampling rate of 44.1 kHz (CARDDELUXE). Prior to the experiment, the two target words were scaled to have equal total root-mean-square (rms) levels and resampled at a rate of 24.414 kHz using MATLAB.

Target words were presented in a continuous background of two-talker speech or speech-shaped noise. The two-talker speech masker consisted of two streams of meaningful speech. Each stream of speech was recorded from a different female talker while reading aloud from popular infant and toddler books; these recordings had a sampling rate of 24.414 kHz and a resolution of 32 bits. The two individual speech streams were manually edited to remove
silent pauses greater than 300 ms, resulting in samples that were 3.5 and 3.1 minutes in duration. Each sample was repeated without discontinuity for 60-minutes, and then the two individual streams were balanced for overall rms level. Next, the two streams were mixed. The spectral envelope of this two-talker masker was used to create the speech-shaped noise masker. The Gaussian noise was transformed into the frequency domain and multiplied by the magnitude spectrum extracted from an equal-duration sample of the two-talker masker. The result was transformed back into the time domain, generating a 95.1-sec sample of noise that could be repeated without discontinuities at the beginning and end of the array.

Custom software (MATLAB) was used to control the selection and presentation of stimuli. The spondee tokens and the masker were mixed (TDT SM3), amplified (Techtron 5507), sent to a headphone buffer (TDT HB6), and presented using a loudspeaker (Monitor Audio, Monitor 4). During testing, the listener was positioned 1 m from the loudspeaker in the sound field of a 7X7 foot, double-walled sound-treated booth. The height and position of the listener’s chair was adjusted so that the stimuli would be presented at approximately 0° azimuth and 0° elevation.

Procedures

Masked speech detection thresholds were estimated using speech produced by a male or a female talker in the presence of a two-female-talker or a speech-shaped noise masker. Each listener was assigned to a single masker type. The target word ‘hotdog’ was used throughout testing for all participants assigned to the two-female-talker masker. The target word ‘playground’ was used throughout testing for all participants assigned to the speech-shaped noise masker. Adults were tested in a single visit to the laboratory. Infants were tested in two separate visits occurring within a 2-wk period. Visits for both age groups were approximately 45 min in length.

A single-interval, observer-based psychophysical procedure was used to test infants (Olsho et al., 1987). The adult procedure was kept as similar as possible to the infant procedure. Infants were tested sitting on their parents’ laps. An assistant sat inside the booth with the parent and infant, manipulating toys in order to keep the infant facing toward the midline. To prevent the assistant and the parent from hearing the tones and influencing the infant’s response, the adults wore noise-isolating earphones (Etymotic mc5) that delivered masking sounds as well as noise-reduction earmuffs (Bilsom Thunder T3). To the listener’s right were two mechanical toys with lights in a dark Plexiglas box. An observer sat outside of the booth and initiated trials when the listener was quiet and facing midline. Adult listeners were tested alone in the booth.

The assigned masker was presented continuously throughout testing at an overall level of 50 dB SPL. Trials were either ‘signals,’ in which the target spondee word was presented, or ‘catch’ trials, in which no target word was presented. The observer did not know which type of trial occurred and was required to decide on the trial type based on the listener’s behavior within 4 s of trial onset. The listener was provided with reinforcement if the observer correctly identified a signal trial. Reinforcement was the activation and illumination of a mechanical toy. The observer was provided with feedback after every trial.

A complete session included two training phases and one testing phase. Target words were presented at a clearly audible level in both training phases, depending on the masker type and age group. The goal of the first training phase was to establish the relationship between the presentation of the target word and the mechanical toy reinforcement. The probability of a signal trial was 0.80, and the probability of a catch trial was 0.20. Listeners were reinforced after each signal trial, regardless of the observer’s response. The first training phase was completed when the observer correctly responded to four of five consecutive trials, including at least one catch trial. The goal of the second training phase was to demonstrate to the listener that he/she was required to respond to signal trials in order to turn on the mechanical toy reinforce. The probability of both signal and catch trials in this phase was 0.50. Reinforcement was only provided to the listener if the observer correctly identified a signal trial. The second training phase was completed when the observer maintained a hit rate of 0.80 or higher, and a catch trial rate of 0.20 or lower.

During the testing phase, thresholds for the target word were measured adaptively using a 2-down, 1-up procedure. The probability of a signal trial was 0.75, and the probability of a catch trial was 0.125. In addition, probe trials were presented with a probability of 0.125. Probe trials were presentations of the target word at the training level. Only signal trials affected the adaptive track. Based on pilot data, the starting level for the target word was about 10 dB higher than the expected threshold value. The initial step size was 4 dB. Eight reversals were obtained, and threshold was based on the last six reversals. Thresholds were only accepted if the response rate to probe trials was 0.60 or higher, and the catch trial rate was 0.40 or lower.
RESULTS

Figure 1 shows masked speech detection thresholds for infants (left panel) and adults (right panel) tested in the two-female-talker masker. The open boxes show the range of performance spanning from the 25th to the 75th percentile for listeners tested in the same-sex condition. Median scores are shown by the horizontal lines inside each box. The 10th and 90th percentiles are shown by the vertical lines. Lower thresholds indicate greater sensitivity. The shaded boxes show the same range of performance for listeners tested in the different-sex condition.

Masked detection thresholds for infants assigned to the same-sex condition (n=8) ranged from 50.0 to 73.5 dB SPL (mean=58.4) and from 45.2 to 75.3 dB SPL (mean=56.2) for infants assigned to the different-sex condition (n=8). Corresponding masked detection thresholds for adults assigned to the same-sex condition (n=8) ranged from 29.0 to 45.9 dB SPL (mean=35.8) and from 24.0 to 33.0 dB SPL (mean = 28.7) for adults assigned to the opposite-sex condition (n=8).

To assess the difference in threshold between listeners tested in the opposite-sex and same-sex conditions, an independent-samples t-test was performed for each age group. No significant difference between infants tested in the opposite-sex and same-sex condition was observed for infants [t_{14}=0.48, p=0.637]. There was a significant difference in threshold in presence of the two-female-talker masker between adults tested in the opposite-sex condition and adults tested in the same-sex condition [t_{14}=3.25, p=0.006]. This significant group effect reflects a difference in average threshold of 7 dB between the adults tested in the same-sex condition (mean=35.8 dB SPL) compared and the adults tested in the opposite-sex condition (mean=28.7 dB SPL).

FIGURE 1. Masked thresholds corresponding to 70.7% correct spondee detection are shown for infants (left panel) and adults (right panel). The open rectangles span the 25th to 75th percentiles for listeners tested in the same-sex condition. The shaded rectangles show this range for listeners testing in the opposite-sex condition. The horizontal line inside each box shows the median. The vertical bars show the 10th and 90th percentiles.

DISCUSSION

While data collection is ongoing, these initial results suggest that infants may derive less benefit than adults from a mismatch in target and masker sex. This would be consistent with an immature ability to use sex-based acoustic differences to segregate auditory streams.

ACKNOWLEDGMENTS

This work was supported by NIH, R01 DC 011038.
REFERENCES


