Available online at www.sciencedirect.com
SCIENCE DIRECT®
COGNITION
ELSEVIER
Cognition 94 (2004) B45–B53
www.elsevier.com/locate/COGNIT

Brief article

Listen to your mother! The role of talker familiarity in infant streaming

Brittan A. Barker[a,*], Rochelle S. Newman[b]

[a]*Department of Otolaryngology-Head and Neck Surgery, University of Iowa, 21200 Pomerantz Family Pavilion, Iowa City, IA 52242, USA*
[b]*Department of Hearing & Speech Sciences and Program in Neuroscience & Cognitive Sciences, University of Maryland, College Park, MD, USA*

Received 15 March 2004; accepted 15 June 2004

Abstract

Little is known about the acoustic cues infants might use to selectively attend to one talker in the presence of background noise. This study examined the role of talker familiarity as a possible cue. Infants either heard their own mothers (maternal-voice condition) or a different infant's mother (novel-voice condition) repeating isolated words while a female distracter voice spoke fluently in the background. Subsequently, infants heard passages produced by the target voice containing either the familiarized, target words or novel words. Infants in the maternal-voice condition listened significantly longer to the passages containing familiar words; infants in the novel-voice condition showed no preference. These results suggest that infants are able to separate the simultaneous speech of two women when one of the voices is highly familiar to them. However, infants seem to find separating the simultaneous speech of two unfamiliar women extremely difficult.
Published by Elsevier B.V.

Keywords: Acoustic cues; Infants; Maternal-voice condition; Streaming; Mother's voice; Talker familiarity

One of the most important tasks facing an infant is learning her native language; isolating the speech stream and the words within that stream is a critical first step in this process. Yet much of the speech that infants hear occurs in the presence of noise. Unless infants can separate the speech from background noise, they are unlikely to learn from

* Corresponding author.
E-mail address: brittan-barker@uiowa.edu (B.A. Barker).

0022-2860/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.cognition.2004.06.001

speech in those settings. In fact, in a survey of families participating in the present study, two-thirds reported that when they speak to their infants, there are *frequently* other members of the household talking simultaneously. Only one parent (of 56) reported that this "almost never" occurred. Thus, to the extent that this parental-report data is accurate, infants seem to be in multi-talker environments quite often.

We know little about how infants succeed at learning language in these situations. Clearly, they do; infants in noisier environments do not appear to be at a serious disadvantage for learning language. However, the cues that infants might rely on to separate the signal from background noise are not yet known. Most research on this problem has focused on adults (see Bregman, 1990, for a review), and suggests that segregating streams of simultaneous speech is a difficult task requiring multiple acoustic cues. Adult listeners find it easier to separate two streams of speech if they come from separate locations in space (Broadbent, 1954; Cherry, 1953; Poulton, 1953), from speakers with different fundamental frequencies or genders (Brokx & Nooteboom, 1982; Darwin & Hukin, 2000) or if there is visual information from a speaker's face to help separate the two voices (Driver, 1996). These findings provide a starting point for understanding simultaneous stream segregation in infants.

Newman and Jusczyk (1996) examined infants' ability to separate speech streams (with a 10 dB S/N) by familiarizing infants with target words spoken by a female talker while a male talker simultaneously read a distracter passage. Following this dual presentation, infants heard test passages in the target voice that contained either words the woman spoke during familiarization or novel words. Infants listened significantly longer to the passages containing the familiarized words. Newman and Jusczyk argued that infants would not have shown a preference for one of the two types of passages unless they had attended to the target voice during familiarization. This finding demonstrates that by 7.5 months of age, infants can selectively attend to one speech stream despite the simultaneous presentation of an additional, competing stream, in some contexts.

Not only did this prior research demonstrate that infants can separate multiple streams of speech, but it began an investigation of the cues that infants might use to do so. Specifically, it showed that infants can selectively attend to a particular target talker when that talker is more intense than the background talker, and when the talkers are perceptually dissimilar (i.e. of different genders). Yet these cues may not always be present in the listening environment. Infants may find themselves in situations in which their caregivers' voices are softer than background voices or in which several women are talking simultaneously. Often the spoken language input an infant receives is neither actually addressed to her nor in close proximity to her, thus decreasing the amplitude of the signal (van de Weijer, 1998). It is likely that infants utilize supplementary acoustic cues to facilitate streaming in these more difficult situations.

Familiarity with the target voice may be an additional cue infants use when streaming. Although infants may frequently be spoken to in noisy environments, they typically are spoken to by a relatively small number of talkers with whom they have had extensive experience. Familiarity with a particular talker might make it easier for an infant to separate that voice from other talkers. Although no research in the adult literature has specifically examined the influence of talker familiarity on streaming, it has been demonstrated to be an important factor in identifying speech in the context of other types

of noise. Magnuson, Yamada, and Nusbaum (1995) presented adults with distorted speech, and found that identification of the degraded speech improved when the talker was a family member of the research participant. Similarly, adults' ability to recognize novel words spoken in noise significantly improves when the words are spoken by a familiar talker (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994). These results suggest that talker familiarity facilitates speech perception for adults, raising the possibility that familiarity with the target acoustic signal may significantly improve infants' streaming performance as well.

The objectives of our study were as follows: (1) to build on the work of Newman and Jusczyk (1996) and determine whether infants can selectively attend to the voice of one woman while a second woman is speaking simultaneously; (2) to determine whether talker familiarity serves as an additional acoustic cue that may compensate for the increased difficulty of streaming two perceptually similar voices (i.e. the voices of two women).

1. Method

To test infants' ability to stream perceptually similar voices in both the presence and absence of a familiar talker, we modeled the experiment after Newman and Jusczyk (1996), using the same methods, words, and sentences; however, the talkers were varied across participants. Half of the infants were assigned to hear their own mothers (maternal-voice condition), and half of the infants were assigned to hear a different infant's mother (novel-voice condition).[1] Each infant in the novel-voice condition was paired with an infant in the maternal-voice condition, and heard that infant's mother's voice. Thus, every infant in a given condition heard a different target voice, but infants across the two groups heard the same voices.

1.1. Participants

A total of 56 full-term infants (25 females), aged 7.45 months (range: 6.57–8.39 months), learning English as their native language, participated in this study.[2] Data from 31 additional infants were excluded from the final analysis for the following reasons: crying/fussiness ($n=20$), ear infections ($n=3$), sleeping ($n=1$), caregiver interference ($n=1$), experimenter/apparatus error ($n=3$), and failing to listen for an average of 1.5 s to each test passage (long enough to hear at least one of the six sentences; $n=3$). Average age of infants in the maternal-voice condition was 7.44 months, that of infants in the novel-voice condition was 7.45 months ($t(54)=0.04$, $p>0.05$).

[1] Infants' mothers were selected as the target voices because the mother's voice is one infants have known longest-even prior to birth (Querlen, Renard, Versyp, Paris-Deline, & Crepin, 1988). Infants also prefer listening to female voices, especially that of their own mother (DeCasper & Fifer, 1980; Kisilevsky et al., 2003).

[2] Participants came from both the University of Iowa and Maryland communities. Both members of each matched-pair were tested in the same location to insure proper matching, and all testing used identical apparatuses. Two of the 56 infants came from bilingual homes, but parents reported that the infants heard more English than any other language.

1.2. Stimuli

Infants were familiarized either with the words *cup* and *dog*, or with the words *feet* and *bike*. During the test phase, they heard four different, six-sentence test passages, each containing one of the four target words. A separate set of target stimuli was recorded for each infant in the maternal-voice condition. Mothers visited the laboratory before their infants' experimental sessions, and recorded the stimuli in the same voice they used "when speaking or reading to your infant". Recordings were made in a double-walled sound booth, then amplified, digitized, and stored on a personal computer. Each mother recorded 15 consecutive repetitions of each target word and four different six-sentence test passages. Although infants only heard two of the four target words in the test session, mothers recorded all four items to prevent differential practicing between the recording and test sessions.

A female native-speaker of American English reading a scientific paper served as the distracter passages. The speaker was instructed to read the passage as if she were speaking to another adult. This recording was taken from a previous study (Newman, 1998).

Both test-passage recordings and target-word recordings were equated for duration and r.m.s. amplitude levels. The average intensity level of the distracter passage was adjusted to 10 dB less than the mothers' target-word recordings. This amplitude-adjusted distracter passage was then mixed with each mother's recordings. The final target stimuli consisted of the infant's mother repeating a target word 15 times and the other female simultaneously reading the distracter passage. The distracter passages and word lists were carefully edited, such that the target words always occurred simultaneously with the distracter voice. Further details of stimulus creation can be found in Newman and Jusczyk (1996).

1.3. Design

The present study employed a yoked-control experimental design (Fig. 1). Infants were assigned to either the maternal-voice or the novel-voice condition. During the familiarization phase, half of the infants in the maternal-voice condition were assigned to hear *cup* and *dog*. The remaining maternal-voice infants heard *bike* and *feet.* During the test phase, each of these infants heard all four of his or her mother's recordings of the six-sentence test passages containing the words *cup*, *dog*, *bike*, or *feet*. Infants heard both the passages containing the target words and those containing the novel words during the test phase; listening times were compared for these two types of passages. Each infant in the novel-voice condition was matched with an infant in the maternal-voice condition, and heard the same recordings as that infant.

1.4. Apparatus

The experiment was conducted in a three-sided booth similar to the one used by Newman and Jusczyk (1996). The booth was constructed of three white pegboard panels. A white light was located in the center of the front panel; a video camera was mounted below the light. A red light was located in the center of each side panel with a loudspeaker mounted behind it, outside of the booth. A computer terminal and a six-button response box were located behind the front wall of the testing booth for the experimenter's access.

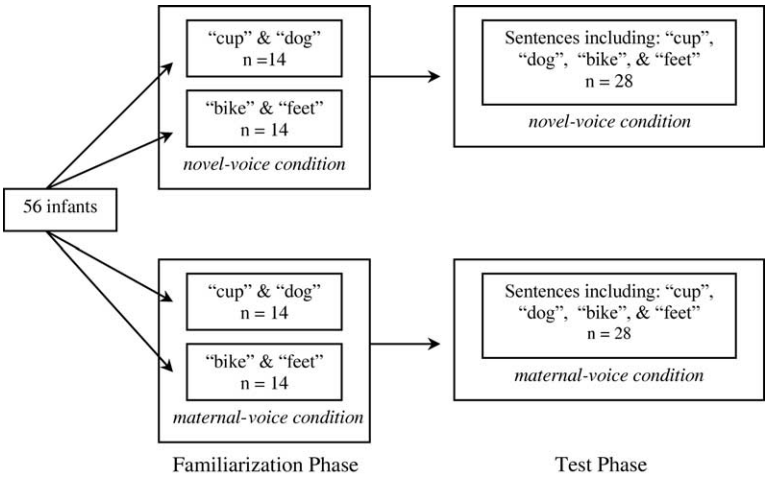

Fig. 1. Schema of yoked-control experimental design for the present study.

1.5. Procedure

A brief hearing screening preceded the experimental session; each experimental session followed the exact testing procedure utilized by Newman and Jusczyk (1996)—a variant of the headturn preference paradigm (Jusczyk and Aslin, 1995; Kemler Nelson et al., 1995).

During the session, the experimenter was seated behind the booth's front panel where she observed the infant and recorded information about the direction and duration of headturns via a response box. Both the mother and experimenter listened to masking music over headphones throughout the experimental session. Interrater reliability was determined by having a separate experimenter re-code videotaped sessions of 20 infants. The correlation on trial-lengths between experimenters was greater than 0.99.

During the familiarization phase, infants heard their two assigned target words (and the simultaneous distracter passages) on alternating trials until they accumulated a minimum of 30 s listening time to each target-word/distracter passage combination. Listening time was determined by the amount of time each infant spent looking at the speech source (the flashing red light).

The test phase began after the familiarization criterion was reached. During this phase, each infant heard all four of the six-sentence test passages in each block, and heard a total of four blocks (16 trials). The computer randomized presentation of the test passages within each block. If an infant began to get fussy, the experimenter used a puppet between each pair of trials on the following block to provide motivation to complete the test phase.

2. Results

Average listening times to the test passages containing familiarized, target words and unfamiliar, novel words were calculated for each infant across the four blocks of trials (Fig. 2).

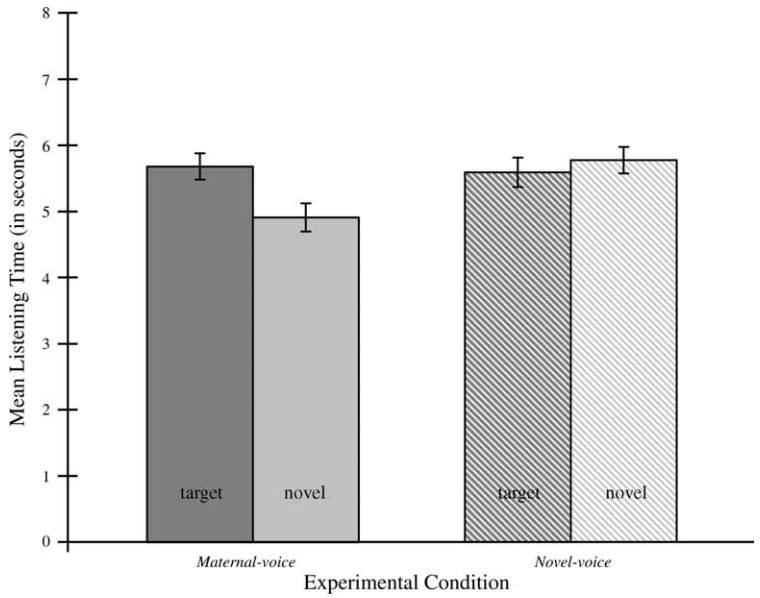

Fig. 2. Mean listening times (in s) and standard errors to the test passages containing the familiar, target words (dark gray bar) and the unfamiliar, novel words (light gray bar). Data from infants listening to their own mother's voice are on the left and that from infants listening to a novel voice are on the right.

A 2×2 ANOVA, with the factors of voice familiarity (mother's voice vs. novel voice) and words (target vs. novel) showed no main effect of voice ($F<1$) or words ($F(1, 27)=2.37, p>.05$); the interaction was significant ($F(1, 27)=4.72, p<.05$). Follow-up t-tests showed that infants in the maternal-voice condition listened significantly longer to passages containing the familiarized, target words ($M=5.68$ s for target-word passages, $M=4.91$ s for novel-word passages, $t(27)=2.67, p<.01$). Eighteen of the 28 infants showed this pattern. This suggests that 7.5-month-old infants can selectively attend to the speech of their mothers while another woman speaks simultaneously.

In contrast, infants in the novel-voice condition did not show a preference for the familiarized, target words despite a 10 dB S/N ($M=5.59$ s for target-word passages, $M=5.78$ s for novel-word passages, $t(27)=-0.63, p>0.05$). Only 10 of the 28 infants in this novel-voice condition listened longer to passages containing the target words. Moreover, this finding is not the result of lower listening times achieved by these infants. There was no main effect of voice, suggesting that infants listening to a novel voice listened as long on average as did infants hearing their own mother; they simply did not show any distinction between target and novel words, thus they did not present any evidence of streaming.

3. Discussion

As adults, we can recognize the difficulty in trying to follow one conversation in the midst of several talkers. Infants presumably find the task of following their caregiver's

speech in a noisy room equally difficult. However, unlike adults, infants have the additional task of acquiring their native language. Unless infants are capable of separating a sound stream into its component sources, they may not be able to benefit from all of the potential language experience to which they are exposed. Infants do seem to have the skills necessary to separate different streams of sound. Prior research demonstrated that infants are capable of attending to a female voice while a male voice speaks simultaneously in the background (Newman & Jusczyk, 1996). The present study extended this finding by examining infants' ability to attend to voices that are more similar. We found that infants were capable of segregating two female voices, but only when one of the two voices was well-known to them.

The present study employed a yoked-control experimental design to demonstrate the importance of talker familiarity as a cue to infant streaming. By having the same voices serve as *familiar* voices to infants in the maternal-voice condition and serve as *unfamiliar* voices to infants in the novel-voice condition, the impact of talker familiarity on streaming was supported.[3] When presented with a familiar voice, infants were capable of separating it from the speech of background talkers, even talkers who share the same gender. Without that cue of familiarity, infants did not show evidence of stream segregation in this listening context.[4] These data suggest that talker familiarity aids infants in separating different streams of speech.

These results also suggest that the perceptual difference between two voices (e.g. a difference in gender) is another important cue to streaming. Both the infants in the novel-voice condition in the present study, and the infants in the Newman and Jusczyk (1996) study were presented with two unfamiliar voices at a 10 dB S/N ratio. The latter infants performed relatively well, showing a preference for the familiarized, target words during the test phase. The infants listening to a novel voice in the current study failed to show such a preference.[5] This implies that infants find the task of separating two simultaneous speech streams more challenging when the voices are perceptually more similar to one another. It appears that the greater perceptual difference between the two voices in the earlier study served as a cue to help the infants separate the speech streams. Thus, this experiment demonstrates that infants use both talker familiarity and perceptual differences as cues to help separate different streams of simultaneous speech.

There are several reasons why familiarity with a voice might have helped infants in the present study. First, the infants had prior experience listening to their mothers' voices. This knowledge of the target voice may help the listener distinguish it from the background voice, thus making streaming easier – much like adults' ability to recognize

[3] Because of the yoked control (i.e. infants in the novel-voice condition), it is unlikely that the impact of speaker familiarity is caused by subtle differences in the mothers' vocal features (e.g. Trainor, Austin, & Desjardins, 2000).

[4] It is unclear whether the infants in the novel-voice condition were *unable* to separate the novel, target voice from the distracter voice or were able to do so but chose to attend to the wrong voice. Although the target voice was created to be more "interesting" (i.e. infant-directed speech) than the distracter voice (i.e. adult-directed speech), we cannot be sure that the infants chose to attend to the target talker. Thus, infants may be able to separate two female voices, even though they showed no evidence of it here.

[5] Although comparisons across studies are hazardous, a statistical comparison does show a significant interaction between word (target vs. novel) and background talker (male vs. female), $F(1, 50) = 16.77$, $p < .0005$.

words in noise or in degraded speech is facilitated when the words are spoken by a familiar talker (Magnuson et al., 1995; Nygaard & Pisoni, 1998; Nygaard et al., 1994). Another possible reason for better performance with a familiar voice is motivational. Reinforcing properties of the maternal voice (DeCasper & Fifer, 1980) may provide additional motivation for infants to devote the processing resources necessary for selective attention. That is, infants may be willing to expend more effort to attend to their mother's voice in the context of noise than to an unknown woman's voice. It is likely that both of these factors played a role in the present study. However, separating motivational and experiential effects, at least with infants, may be impossible because familiarity tends to be inextricably linked with motivational biases (Hunter & Ames, 1988; Rose, Gottfried, Melloy-Carminar, & Bridger, 1982).

In conclusion, the present results begin to flesh out the cues infants use in simultaneous streaming. Voice familiarity and similarity are both important cues that infants can use to help them attend selectively to a particular voice in the context of a multi-talker environment. Continuing to examine the role different cues play in infants' stream segregation should ultimately allow us to determine what limits there may be on infants' ability to learn language from noisy environments.

Acknowledgements

This research was supported by Research Grants HD37822-01 from NICHD and BCS 99-07849 from NSF to the University of Iowa, and by a University of Iowa CIFRE award to the second author. We thank James Sawusch for development of the infant subject-running system, Lloyd Frei and Keith Miller for design and implementation of the test booth, Ruth Tincoff for stimulus recording, Sheryl Clouse for re-coding, and Lisa Oakes for commentary on a draft of this manuscript. We also thank Jodie Plumert and the late Peter Jusczyk for many helpful discussions. Finally, we extend our gratitude to the families who volunteered their voices and time for these studies. Some of these results were presented previously at the 24th annual Boston University Conference on Language Development, and may be found in the Proceedings from that conference.

References

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

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, *47*, 191–196.

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, *10*, 23–36.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975–979.

Darwin, C. J., & Hukin, R. W. (2000). Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *Journal of the Acoustical Society of America*, *107*, 970–977.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*, 1174–1176.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, *381*, 66–68.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, *5*, 69–95.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *28*, 1–23.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). The Headturn Preference Procedure for testing auditory perception. *Infant Behavior and Development*, *18*, 111–116.

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. *Psychological Science*, *14*, 220–224.

Magnuson, J. S., Yamada, R. A., & Nusbaum, H. C. (1995). The effects of familiarity with a voice on speech perception. *Proceedings of the 1995 Spring Meeting of the Acoustical Society of Japan* , 391–392.

Newman, R.S. (1998). *Infants' ability to stream two female voices*. Unpublished manuscript.

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

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*, 355–376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*, 42–46.

Poulton, E. C. (1953). Two-channel listening. *Journal of Experimental Psychology*, *46*, 91–96.

Querlen, D., Renard, X., Versyp, F., Paris-Deline, L., & Crepin, G. (1988). Fetal hearing. *Journal of Obstetrics and Gynecology and Reproductive Biology*, *29*, 191–212.

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology*, *18*, 704–713.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, *11*, 188–195.

van de Weijer, J. (1998). *Language input for word discovery*. Wageningen: Ponsen & Loiiken, b.v. University of Nijmegen: Max Planck Institute, The Netherlands.