The development of speech segmentation abilities

Rochelle Newman
University of Maryland

Jane Tsay
National Chung Cheng University

Peter Jusczyk
Johns Hopkins University

In order to learn the words in their native language, children must have the ability to segment fluent speech into its component words. Parents rarely speak words in isolation (Aslin, 1993) and word boundaries in fluent speech are not clearly marked by pauses or other obvious cues to segmentation (Cole & Jakimik, 1980). Indeed, adult learners of a foreign language often report severe difficulties finding the boundaries between words when they first start listening to the second language (Dejean de la Batie & Bradley, 1995; Koster, 1996), and infants learning their first language are in a similar position. Thus, learning to segment speech is a necessary first step to acquiring a vocabulary.

Previous work has demonstrated that the ability to segment fluent speech develops between 6 and 7.5 months of age (Jusczyk & Aslin, 1995). This learning appears to be language specific; for example, infants of this age do not appear to segment fluent speech in a novel language (Tsay & Jusczyk, unpublished).

This raises the question of how much exposure children actually need before they begin to acquire this segmentation ability. The present study is a first step at addressing this issue. Children were presented with approximately 5 hours of videotaped exposure to Mandarin Chinese prior to visiting the laboratory. They then were tested in the same procedure as that of Tsay and Jusczyk, using stimuli in Mandarin. If this amount of exposure is sufficient for infants to begin acquiring segmentation skills, we would expect that infants in the present study would perform better at the segmentation task than did infants without such exposure (those in the prior study by Tsay and Jusczyk).

Methods

Participants

Eleven infants participated (8 males, 3 females; average age at test 7.7 months (range 7.0 – 8.2). Data from an additional infant was excluded as a result of experimenter error.

Stimuli

Stimuli were taken from Tsay and Jusczyk (unpublished). A female native speaker of Mandarin Chinese from Taiwan recorded 15 repetitions of each of four target words (here written in English equivalences, with tone in brackets): *bei* [H], *dan* [HL], *tian* [H], and *tou* [LH], meaning “cup”, “egg”, “sky/day,” and “head”, respectively. She also recorded four different 6-sentence test passages, in which each sentence included the

target word (see appendix 1 for passages). The target words appeared in a variety of sentential locations in the 6 sentences, so as to prevent their being spoken with similar intonation contours in the different sentences.

The recordings were made in a sound-attenuated room, using a Shure microphone. They were amplified, low-pass filtered at 48 kHz, digitized via a 12-bit, analog-to-digital converter at a 10 kHz sampling rate and stored on a VAX Station model 3176 computer.

The videotape used in this study included a number of short cartoons spoken in Mandarin Chinese. The cartoons had bright colors and animation, and were expected to be appealing to young children. None of the cartoons included any of the target words.

Apparatus.

A PDP 11/73 controlled the presentation of the stimuli and recorded the observer's coding of the infant's responses. A 12-bit D/A converter was used to recreate the audio signal. The resulting audio signal was then filtered at 4.8 kHz, amplified, and played from two loudspeakers mounted on the side walls of the testing booth.

The experiment took place in a 3-sided test booth constructed out of pegboard panels (4' x 6'). An experimenter was located behind the front wall of the booth, and looked through the existing holes in the pegboard to observe the infant. The remainder of the pegboard panels were backed with white cardboard so that the infant could not see movements behind the panels. There was a green light in the center of the front panel, and a 5 cm hole for the lens of a video camera. The video camera was used to provide a permanent record of each session. The two side walls each had a red light and a loudspeaker located in the center of the panel. A white curtain was suspended from the ceiling and prevented the infant from seeing over the top of the booth. A computer terminal and 6-button response box were located behind the front wall of the booth. The experimenter pressed buttons to signal the computer to start and stop the flashing center and side lights.

Test procedure

Experimenters visited children in their homes, and played the videotape for them. They then left the videotape with the parents, with the instructions to play the videotape a 2nd time that day, and then twice per day for an additional 4 days. One week following the last video showing, parents and infants visited the lab for the test session.

Test sessions used the Headturn Preference Paradigm (Kemler Nelson, Jusczyk, Mandel, Myers, Turk, & Gerken, 1995), and were modeled after prior work by Jusczyk and Aslin (1995). During the test session, infants sat on their caregiver's lap in the test booth. During the familiarization phase, infants heard two of the target words on alternating trials until they accumulated at least 30 s. of listening time to each word. (Infants heard either *tou* "head" and *bei* "cup" during the familiarization phase, or *dan* "egg" and *tian* "sky, day".) Listening time was assessed by the amount of time the infant spent looking at the "source" of the sound (the flashing light). The loudspeaker used for each word was randomly varied from trial to trial. A trial was terminated when the infant looked away for more than 2 seconds or when the end of the word list was reached. Information about the direction and duration of headturns and the total trial duration were stored in a data file on the computer.

The test phase began immediately after the familiarization criteria was reached. The test trials were blocked in groups of four so that each test passage occurred once in a given block, although the order of the four passages within each block was randomized. Each infant was tested on a total of 4 blocks, for a total of 16 test trials.

Both familiarization and test trials began by blinking the green light in the center of the front panel. Once the infant had oriented in that direction, the light was turned off and one of the two red lights began to flash. Once the infant had oriented towards that light, the stimulus for that trial (regardless of whether it was a familiarization stimulus or a test stimulus) began to play from the loudspeaker on the same side. The stimulus continued to play until its completion, or until the infant had looked away for 2 consecutive seconds, whichever came first. Any time the infant spent looking away (whether it was 2 seconds or less) was not included when measuring the total listening time. The red light continued to flash for the duration of the entire trial.

The experimenter behind the center panel pressed a button on the response box whenever the infant looked at or away from the flashing light. The experimenter was not told ahead of time which items were used in the familiarization phase, and both the experimenter and the caregiver listened to masking music over headphones (SONY MDR-V600). Reliability checks between the experimenter and observers of the video tapes of each session are high, with correlations ranging from .92 to .96 (see Kemler Nelson et al., 1995), although it is important to note that these estimates of reliability refer to the procedure in general, not to these specific experiments.

Results and Discussion

Mean listening times to the four different test passages were calculated for each infant across the four blocks of trials, and then these times were averaged for the test passages containing the familiarized words and those containing the unfamiliarized words. Average listening to the familiar words was 5.9 seconds (std. dev. 3.09 sec); average listening to the unfamiliar words was 6.0 seconds (std. dev. 2.39 sec). These values did not differ significantly $t(10)=-0.16$, $p>.05$). This null result suggests that infants were not able to segment the words from the fluent speech passages, despite their familiarization with the language.

Acknowledgements

This study was conducted with support from Research Grant from NICHD (HD-15795) to the third author, an NSF Predoctoral Fellowship to the first author, and a postdoctoral trainee grant from NIDCD (5T32DC00036-02) which supported the second author. We thank Ann Marie Juscyzk for her assistance in testing participants. The first author may be contacted at the Department of Speech and Hearing Sciences, University of Maryland, College Park, MD, 20742, rnewman@hesp.umd.edu.

References

Aslin, R. N. (1993). Segmentation of fluent speech into words: Learning models and the role of maternal input. In B. d. Boysson-Bardies, S. d. Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life* (pp. 305-315). Dordrecht, The Netherlands: Kluwer Academic.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech.* (pp. 133-163). Hillsdale, NJ: Erlbaum.

Dejean de la Batie, B., & Bradley, D. C. (1995). Resolving word boundaries in spoken French: Native and non-native strategies. *Applied Psycholinguistics, 16*, 59-81

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

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.

Koster, C. J. (1996). Some experimental tasks in the study of word recognition. In P. Jordens & J. Lalleman (Eds.), *Investigating Second Language Acquisition* (pp. 117-134). Berlin: Mouton de Gruyter.

Tsay, J. & Jusczyk, P. W. (unpub.) Detection of words in fluent Chinese by English-acquiring and Chinese-acquiring infants. Unpublished manuscript.

Appendix
Four test passages

Meige ren de ***tou*** dou bu yiyang. "Everybody's head is different."
Tou keyi cangzai maozi dixia. "A head can be hidden inside a hat."
Wo de hao pengyou shi Da-***tou***. "My best friend is Big Head."
Shizi de ***tou*** hen qiguai. "The head of a lion is very weird."
Tou buyao shenchu che wai. "Don't stick your head outside of the car."
Linju de nanhai jiao Xiao-***tou***. "The neighbor's boy is called Small Head."

Dan fangzai bingxiang li. "Eggs are put in the refrigerator."
Wo mai da ***dan***-gao gei ni. "I bought a big cake (lit. egg-cake) for you."
Muji xiale haoduo ***dan***. "The hen laid many eggs."
Youxie ren bu ai chi ***dan***. "Some people don't like eating eggs."
Xiaohua ba ***dan*** dapole. "Xiaohua broke the eggs."
Dan gundao zhuo-zi dixia qu le. "Eggs rolled down under the table."

Tian shang you ji duo baiyun. "There are some clouds in the sky."
Ta tangzai dishang kan lan-***tian***. "He is lying on the ground looking at the sky."
Women xingqi-***tian*** qu kao rou. "We had a barbecue on Sunday (lit. week-day)."
Tian kong feichang de yin'an. "The sky is very dark."
Dajia dou xihuan qing ***tian***. "Everybody likes clear days."
Xiayu ***tian*** bu neng qu dongwu yuan. "One can't go to the zoo in raining days."

Milaoshu zai zhao cha-***bei***. "Mickey Mouse is looking for a teacup."
Zhe ge xiao ***bei***-zi hen piaoliang. "This little cup is very pretty."
Bei-zi li you guozhi. "There is juice in the cup."
Bei di zuozhe yi zhi qingwa. "There is a frog sitting in the bottom of the cup."
Ta ba yi da ***bei*** niunai heguangle. "He drank up a big cup of milk."
Wo song ni yi ge boli ***bei***. "I'll give you a glass cup."