



The cocktail party effect: Infants' use of visual information in speech segmentation



George J. Hollich¹, Peter W. Jusczyk¹, & Rochelle Newman²

¹ Johns Hopkins University ² University of Maryland

Abstract

7.5-month-olds' abilities to use visual/auditory correlations to segment a complex speech stream were studied using the head-turn preference procedure (HPP) following video familiarization. When two blended voices of equal loudness were presented, infants could use visual correspondences to reliably recognize words in the video, even in the case of a moving oscilloscope pattern. In contrast, in cases where a static picture was presented, infants did not show an ability to segment the speech stream at this signal to noise ratio.

Introduction

- One of the first tasks faced by an infant is to **segment words** in the speech stream.
- Words in speech are not separated with spaces.
- Less than 7% of the speech directed at infants consists of isolated words.
- Even more problematic is the fact that the environment around the infant can be quite noisy.

Are infants able follow the voice of one talker in the midst of competition from other voices and segment that talker's speech stream?

Using the HPP alone, Newman and Jusczyk (1996) found that infants **were** able to segment individual words from a blended stimulus created by combining the fluent speech of two voices.

However, infant only showed evidence for this if the target voice was **at least 10 dB louder** than the distracting stimuli.

The current studies examine whether infants can use **auditory-visual correspondences** to boost their segmentation abilities.

Procedure

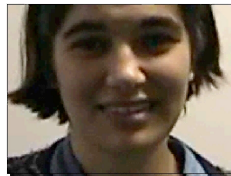
- 1) Infants were familiarized with a **video** display (either a female talker or an oscilloscope display). This was accompanied by an audio track of a **blended stimulus**, consisting of the female talker and a male talker. The female talker's passages were recorded in **infant directed speech** and contained **two target words**: either cup/dog, or bike/feet. The male talker's passage was a selection from the methods section of the original Newman & Jusczyk (1996) paper and was presented in adult directed speech.
- 2) Infants' memory for tokens of the target words was tested using the Headturn Preference Procedure. Tokens were presented by the female talker at 72dB.

Study 1: Hypothesis

Infants can use the visual synchronization between a talker's face and the speech stream to focus on that stream in a blended stimulus.

Study 1: Familiarization (40 sec)

Display

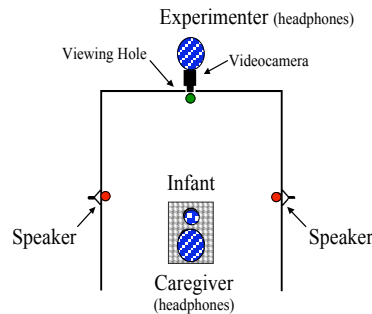


Audio

Female: "The **cup** was bright and shiny. A clown drank from the red **cup**. The other one picked up the big **cup**..."

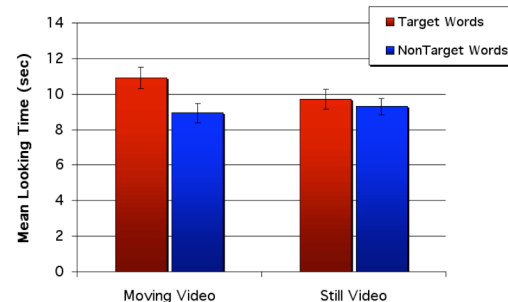
Male: "...on alternating trials before they had accumulated..."

Test: The head-turn preference procedure.



- Tested on two pairs of words: **cup/dog** and **bike/feet**
- **12 trials:** One word per trial, 15 exemplars.

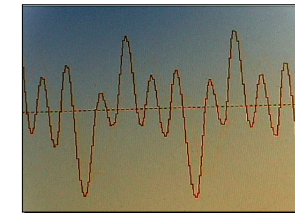
Study 1: Results (n = 30 and 30)



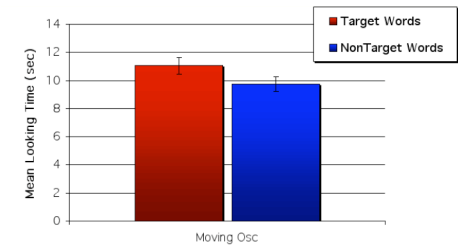
Study 2: Hypothesis

Infants can use the visual synchronization between an oscilloscope and the speech stream to focus on that stream in a blended stimulus.

Study 2: Familiarization (40 sec)



Study 2: Results (n= 26)



Conclusions

- Infants were significantly better at segmenting the words in the moving video condition than in the still video condition at a 0dB S/N ratio.
- Infants can use **visual correspondence to segment speech** even in a noisy environment.
- They can also use this visual information without prior experience, and without the need for domain-specific mechanisms.

INFANTS CAN USE WHAT THEY SEE TO HEAR BETTER.

Correspondence:

George J. Hollich
(ghollich@yahoo.com)
Department of Psychology,
Johns Hopkins University
Baltimore, MD 21218

This research was sponsored by NICHD Grant # 15795 and from a Senior Scientist award from NIMH (#01490) to P.W.J.