3010. Learned cross-modal integration of novel visual cues with auditory speech

Joseph D.W. Stephens & Lori L. Holt

Psychology Dept., Carnegie Mellon University, and Center for the Neural Basis of Cognition

CNBC Center for the Neural Basis of Cognition

Carnegie Mellon

❶ Introduction

Multimodal perception & experience

Experience affects multimodal perception

Sensory modalities can recalibrate each other (e.g., Bertelson et al., 2003; Recanzone, 1998; Atkins et al., 2001)

Tuning of perceptual categories within modalities affects multimodal responses (Massaro et al., 1993)

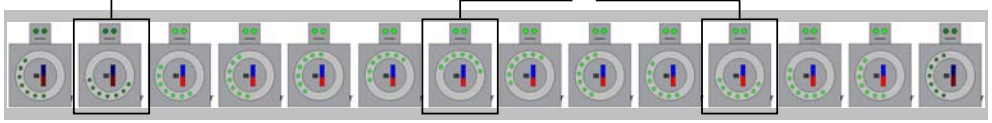
How does information come to be combined across modalities?

Most experiments deal with established multimodal effects

We attempted to train individuals to combine auditory speech and completely novel visual cues

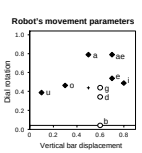
❷ Novel visual cues

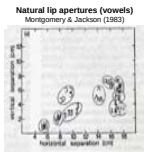
/a b a/



Stimuli

A "speech robot" contained moving parts synchronized with acoustic VCV tokens.



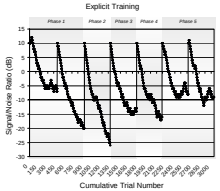


The robot's parts moved to positions defined by phonetic categories. The combinations of cues were derived from natural mouth positions, but bore no resemblance to them (see figures).

Training

Participants were trained to identify consonants across multiple sessions.

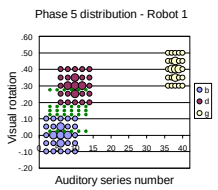
In an **explicit** task, participants identified the consonant spoken by the robot on each trial. Correct responses resulted in increased acoustic noise. In an **implicit** task, participants watched AV stimuli while monitoring for "robot malfunctions." The tasks alternated daily.

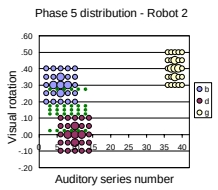


The overall length of training depended on performance in the explicit task.

Examples of stimuli may be viewed at: http://www.andrew.cmu.edu/~jds2/robot.html

The distributions of AV combinations broadened gradually throughout training. Two groups of participants were trained on different distributions. One was analogous to natural AV speech (Robot 1); in the other (Robot 2), bimodal categories were rearranged relative to each other.

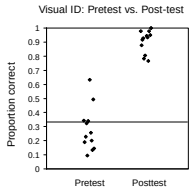




❸ Results

Learning of visual cues

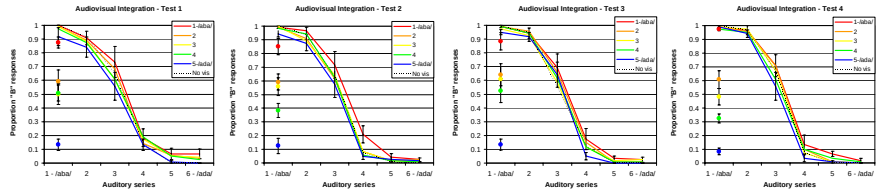
After training, participants were able to identify consonants based on visual cues alone.



Use of visual cues

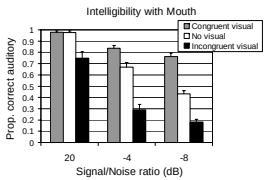
The influence of visual information on consonant identification developed over the course of training

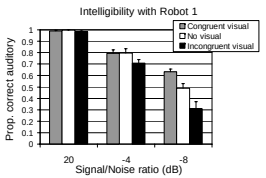
Between training phases participants identified stimuli that consisted of an auditory series ranging from /aba/ to /ada/, paired with robot video ranging from /aba/ to /ada/.



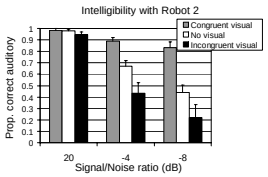
Participants learned to identify visual stimuli early in training, but the effect on bimodal perception was not reliable until the second test. As training progressed, the effect of visual cues became more concentrated around the most ambiguous auditory stimuli.

After training, visual cues from the robot improved intelligibility of consonants in noise





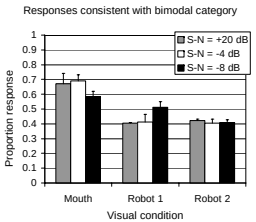
When noise was added to auditory stimuli, congruent visual cues from the robot improved identification, and incongruent cues impaired identification, relative to a static picture of the robot.



Integration of visual cues?

When A and V cues were mismatched, each AV combination still fell within one of the trained 2-dimensional distributions.

However, these bimodal categories were not perceived as reliably with the robot as with cues from a speaker's mouth.



❹ Conclusions

1. **Non-gestural visual speech cues can be learned and used in speech identification**
   - Participants successfully learned novel visual cues for consonants
   - Visual cues affected intelligibility; potential applications
2. **Current effects might not involve same perceptual mechanisms as natural AV integration**
   - Sub-optimal use of visual cues in bimodal perception
   - Weak evidence that modalities are combined on individual trials
   - Limited experience with full distributions; further research necessary

Future research

Further develop training methods

Apply computational models to data

Evaluate integration using additional behavioral tasks

Study usefulness of novel visual cues for improving intelligibility

References

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). *Vision Research, 41*, 449-461.

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). *Psychological Science, 14*, 592-597.

Massaro, D. W., Cohen, M. M., & Gesi, A. T. (1993). *Percept. Psychophys., 53*, 549-562.

Recanzone, G. (1998). *Proc. Natl. Acad. Sci. U. S. A., 95*, 869-875.

This research is supported by Kirchstein-NRSA 1 F31 DC007284-01 and by Riksbankens Jubileumsfond, the Bank of Sweden Tercentenary Foundation