

1pSC12. Experience-driven effects of visual cues in speech perception

Joseph D.W. Stephens & Lori L. Holt

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



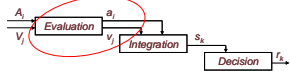
1 Introduction

Visual cues contribute to speech perception (e.g., McGurk & MacDonald, 1976)

How does learning affect audiovisual information integration?

Direct realism (e.g., Fowler & Deka, 1991) claims that learning is unnecessary

FLMP (e.g., Massaro, 1998) assumes that learning affects *evaluation* but not *integration*



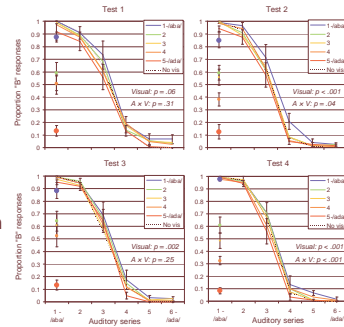
However, prior experiments have used cues that were already well-learned

Here, participants were trained on completely novel visual speech cues

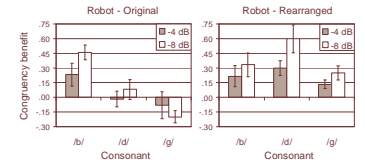
3 Results

Learned visual cues influenced speech identification

Between phases, participants were tested on identification of AV combinations across /b/-/d/ series.



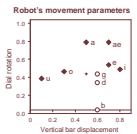
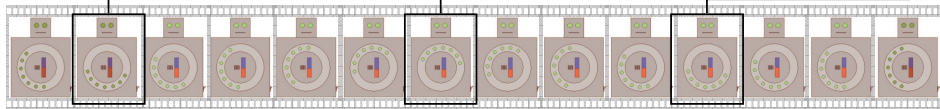
Participants learned to identify visual stimuli early in training, but the effect on bimodal perception was not reliable until the second test. As training progressed, the effect of visual cues became more concentrated around the most ambiguous auditory stimuli.



When noise was added to auditory stimuli, congruent visual cues from the robot improved identification. The nature of this effect depended on the relative arrangement of AV distributions.

2 Novel visual cues

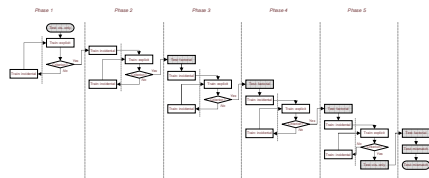
A "speech robot" contained moving parts synchronized with acoustic VCV tokens. /aba/



The positions of the robot's moving parts corresponded to phonetic categories. The moving parts bore no resemblance to human speech articulators.

Training

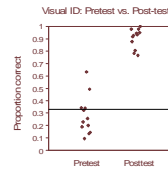
Participants were trained across multiple sessions on combinations of novel visual cues with auditory speech



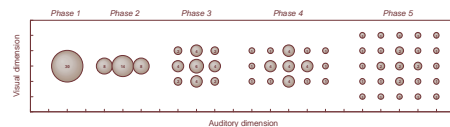
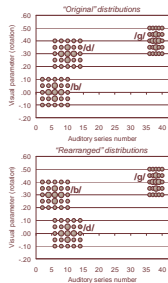
The overall length of training depended on attainment of performance criteria.

Learning

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



Two groups of participants were trained on different distributions of AV combinations. One was analogous to natural AV speech; in the other, bimodal categories were rearranged relative to each other.



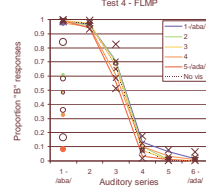
The distributions of AV combinations broadened gradually throughout training.

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

Computational models of integration

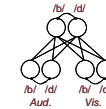
Fuzzy Logical Model of Perception

FLMP predicts optimal AV integration (Massaro, 1998)

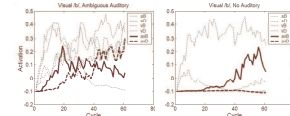


Poor fit to current data; integration not optimal.

Modified Stochastic Interactive Activation Model

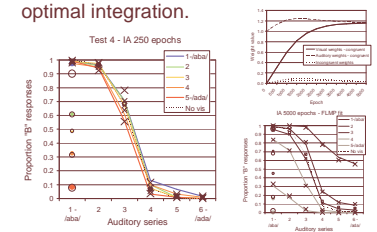


Stochastic IA models can make same predictions as FLMP (McClelland, 1991)



Here, a stochastic IA model was modified so that noise was scaled according to unit activation.

Combined with a simple learning algorithm, the scaled-noise model gave a better fit to the current data and predicted that further learning would lead to more optimal integration.



4 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. Newly-learned integration differs from natural AV integration

Sub-optimal use of visual cues in bimodal perception

3. Integration mechanisms might change with experience

Scaled-noise, stochastic IA model can bridge integration patterns through learning

Future research

Further develop training methods

Evaluate integration using additional behavioral tasks

Study usefulness of novel visual cues for improving intelligibility

References

Fowler, C. A., & Deka, D. A. (1991). *J. Exp. Psychol. Hum. Percept. Perform.*, 17, 818-828.
 Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
 McClelland, J. L. (1991). *Cognitive Psychology*, 23, 1-44.
 McGurk, H., & MacDonald, J. (1976). *Nature*, 264, 746-748.