# Visual influences on interactive speech alignment

Simon Fraser University
Language and Brain Lab

Dias et Rosenblum, 2011

# Overview

- Focus: Speech Accommodation Hypothesis

- Methodology: Conversational AV Study

- Core Ideas:

    1. Theoretical background

    2. Research Objectives

    3. Experimental design and results

# Theories

- **Speech Alignment**: unconscious behavioural tendency to produce speech that shares characteristics with perceived speech

  - **Speech Accommodation Hypothesis**: alignment done to modulate social distance

  - **Learning**: imitation, the natural way of learning a language

  - Not just during interaction, recordings also have effects
Imitation could be unconscious

# Speech Alignment

- Include but not limited to

  - Physical behaviours: posture, gesture, head motion, facial expressions

    - Mouth movements

  - Linguistic features: utterance length, response latency, pausing, info density, self-disclosure

    - Acoustical characteristics: rate, intensity, etc.

BG

# This paper

- This work

  - **Whether** seeing a talker (AV) **enhances alignment** over just hearing a talker (AO)

  - Two groups of participants, each completing interactive tasks

    - AV group: pairs of participants can see and hear each other

    - AO group: pairs of participants can only hear each other
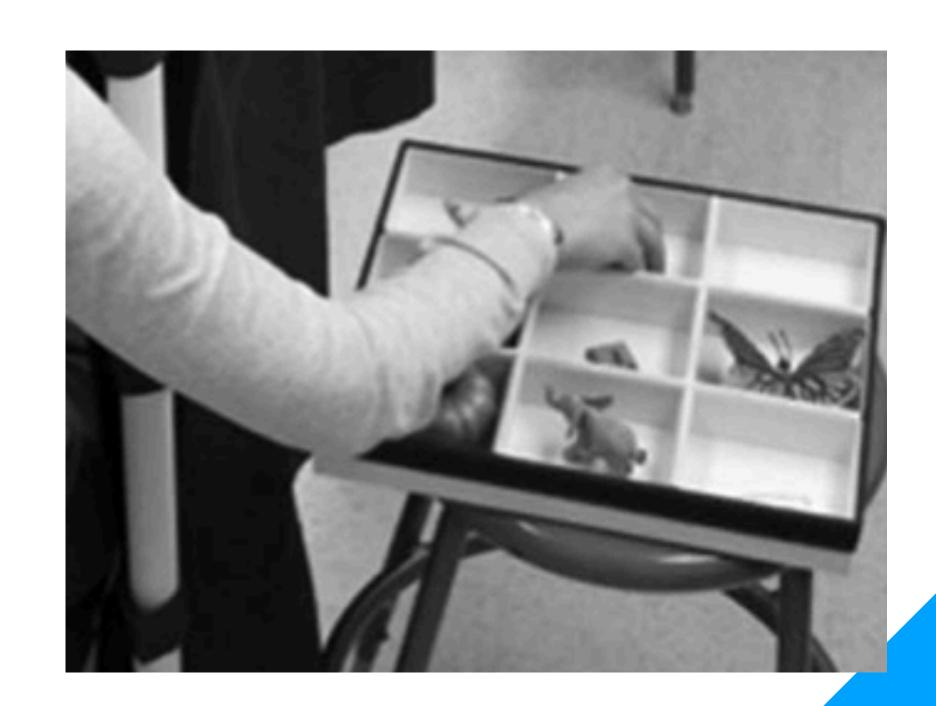
**Theories**

# Experimental Settings

- Pairs of participants

  - AV has full unobstructed view

  - AO has curtain obstructing view

  - Sound unobstructed

# Experimental Settings

- Object search/arrange task

  - Each participant has 3x3 grid, containing randomly placed objects

  - Each participant receives the same objects

  - Neither participants can see the objects, but can use hand to feel

  - Participants need to communicate and match arrangements of the grid



Expt.

# Experimental Settings

- Interlocutors: all female undergraduate subjects with no prior social interaction

- Evaluation: Pre-test and Post-test speech comparison, using human rater

  - nine different carrier phrases which included the nine token items used in object search/arrangement tas

  - token items are placed in the centre of the sentence

- Total length: pre-test, test, post-test combined takes about 1hour

Expt.

# Alignment Assessment

- Alignment assessment: native speakers using AXB paradigm

  - Group 1: compare interlocutor A's last-utterance with partner B's pre-test and last-utterance (Pretest-to-interaction)

  - Group 2: compare interlocutor A's post-test with partner B's pre-test and post-test (Pretest-to-posttest)

  - Predictions

    - if alignment takes place, A&B will sound more similar during and shortly-after interaction

    - visual information should enhance alignment

Expt.

# Pretest-to-interaction

- Raters of pretest-to-interaction tokens detected levels of alignment between audio-only (AO) interlocutors significantly above chance

- Ratings of audiovisual (AV) interlocutors also showed detectable levels of alignment significantly above chance

- Ratings for AV tokens show significantly greater alignment than those for AO alignment

# Conclusions

- auditory information alone is sufficient for speech alignment to occur

- Speech alignment significant for AV

  - Pretest to interaction-utterances: AV greater than AO

  - Pretest to posttest utterances: AV greater than AO

- Both AV and AO

  - Alignment greater for the pretest-to-interaction than for pretest-to-posttest