Acoustic-to-Articulatory Inversion based on Dynamic MRI

Zeynab Raeesy (zeynab.raeesy@phon.ox.ac.uk)

Phonetics Laboratory – University of Oxford

Introduction and Motivation

Understanding the underlying speech production process:

➢ Direct Observation

Problem: Physiological interventions are inevitable, affecting the naturalness of speech and speech production

➢ Modeling the speech production system using more accessible features of speech for obtaining articulatory information

Acoustic-to-articulatory Inversion:
The problem of finding a mapping between the acoustic and articulatory spaces having the acoustic data parameters.

Direct function

Inversion
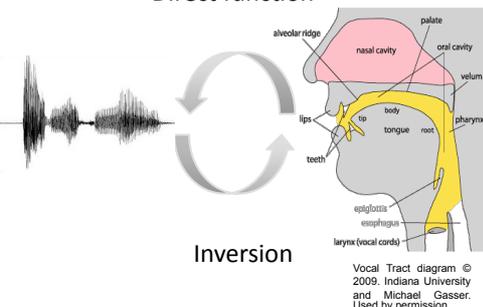
Vocal Tract diagram © 2009. Indiana University and Michael Gasser. Used by permission.

Motivations:

➢Scientific purposes:
More comprehensive understanding of human speech

➢Practical applications:
Training technologies for people with speaking difficulties and foreign learners of a language, aiding systems for hearing impaired people, speech recognition and synthesis, and speech visualization for gaming and animation.

Dynamic MRI Database

MRI data set of Phonetics laboratory at university of Oxford [1]:

➢Dynamic MRI image sequences
➢20 native British English speakers
➢Phrases were chosen to be short (4-6 syllables) for repeatability
➢Each phrase was repeated 20 times
➢Speakers talked to a metronome beat which was set individually for each speaker
➢Speakers' voice was recorded simultaneously with a noise-cancelling non-magnetic microphone
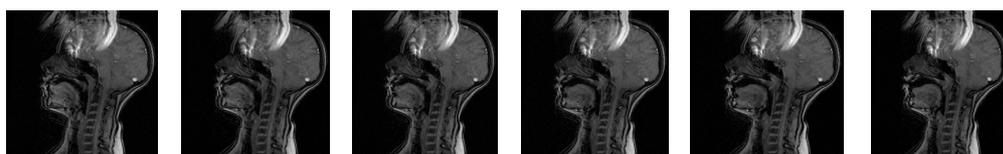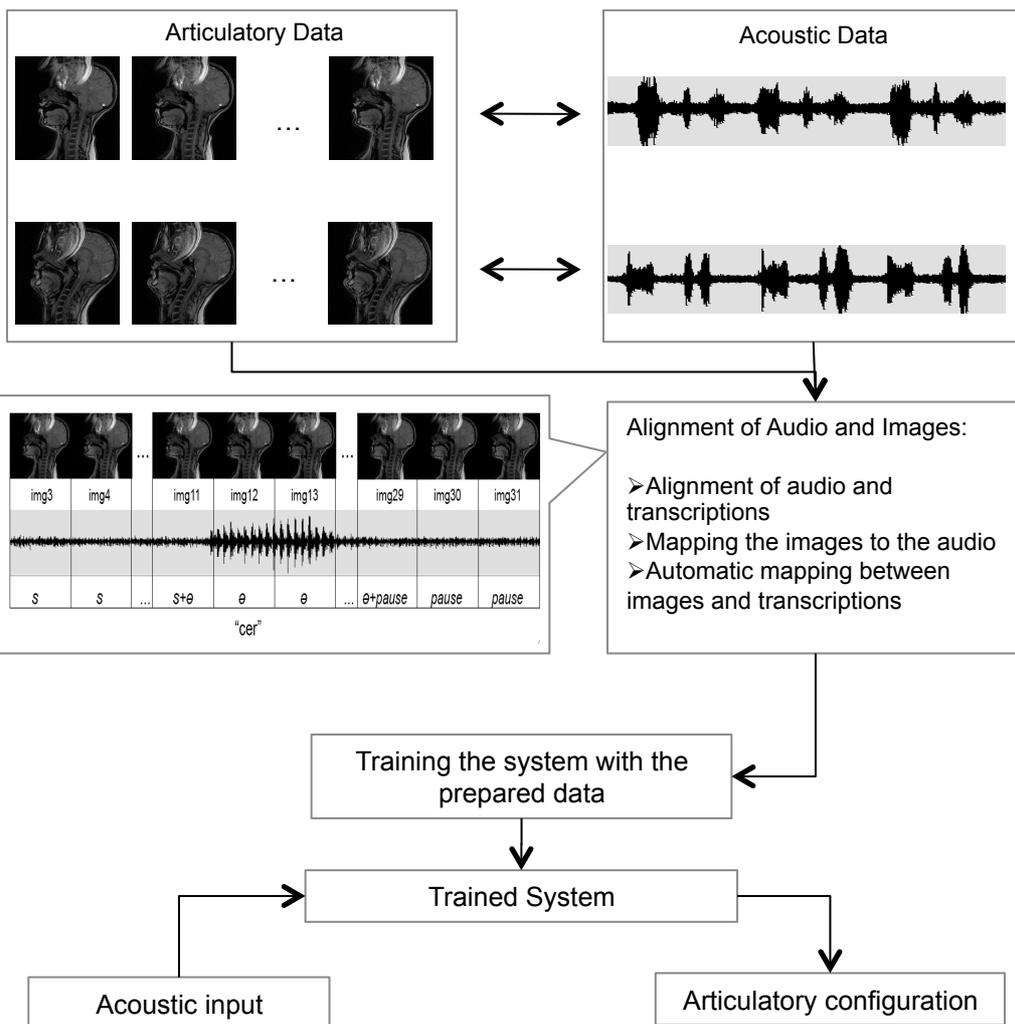➢MRI scanner noise was subtracted out of the audio recording (by G. Kochanski)

Image sequence roughly corresponding to the articulation of /fɔ/ by a female speaker.

A Data-Driven Inversion Technique

Articulatory Data

Acoustic Data

img3 img4 ... img11 img12 img13 ... img29 img30 img31

s   s   ...   s+ə   ə   ə   ...   ə+pause   pause   pause

"cer"

Alignment of Audio and Images:

➢Alignment of audio and transcriptions
➢Mapping the images to the audio
➢Automatic mapping between images and transcriptions

Training the system with the prepared data

Trained System

Acoustic input

Articulatory configuration

Typical Challenges

➢Quality of the audio data recorded inside the MRI scanner
  o Noisy data and narrow microphone bandwidth
➢Alignment of speech signal with the obtained image sequence
  o Image sequence does not correspond to any individual production (dynamic MRI)
➢Variability in articulatory data
  o Inter-speaker and intra-speaker variability

Citations

[1] Alvey, C., C. Orphanidou, J. Coleman, A. McIntyre, S. Golding and G. Kochanski. 2008. Image quality in non-gated vs. gated reconstruction of tongue motion using Magnetic Resonance Imaging: A comparison using automated image processing. International Journal of Computer Assisted Radiography and Surgery, 3 (5), 457- 464.