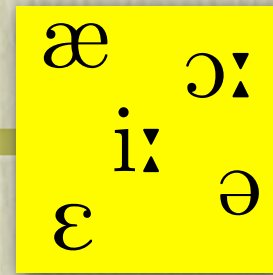
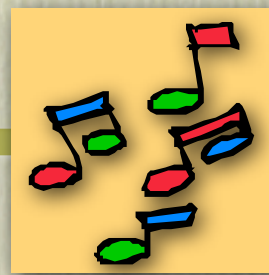


Speech as Timbre-based Melody

-- What in parents' voices do infants imitate acoustically? --



¹Nobuaki Minematsu & ²Tazuko Nishimura

¹Graduate School of Engineering

²Graduate School of Medicine

The University of Tokyo



Two kinds of vocal imitation

Vocal imitation by myna birds

Imitation of voices

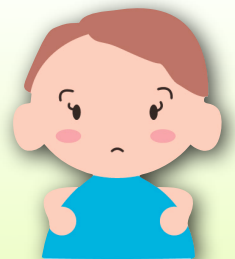
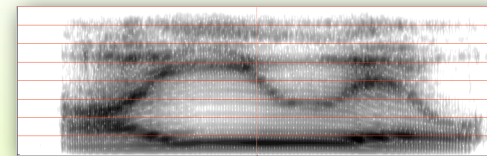
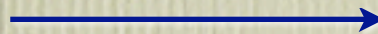
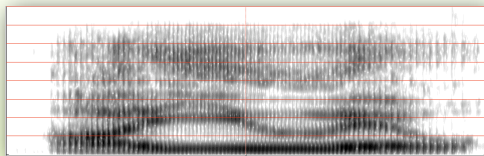
- Myna birds imitate the sounds of cars, animals, and human voices.
- Hearing an adept myna bird, one can guess its owner (Miyamoto'95).



Vocal imitation by young children

Imitation of **some aspects** of voices

- Children don't try to produce their fathers' voices.
- They perceive equivalence between their utterances and fathers' ones.



Two kinds of vocal imitation

Vocal imitation by myna birds



Imitation of voices

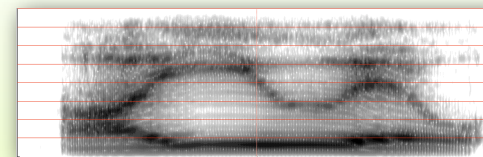
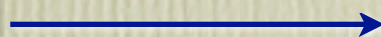
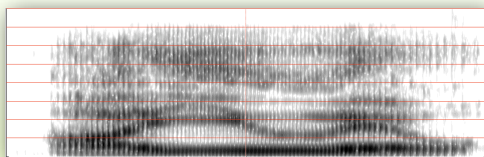
- Myna birds imitate the sounds of cars, animals, and human voices.
- Hearing an adept myna bird, one can guess its owner (Miyamoto'95).

Vocal imitation by young children



Imitation of **some aspects** of voices

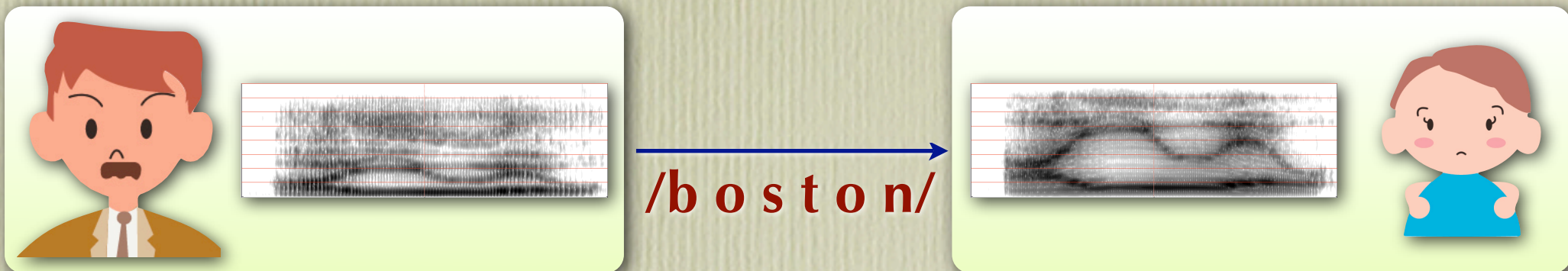
- Children don't try to produce their fathers' voices.
- They perceive equivalence between their utterances and fathers' ones.



What in voices do infants imitate acoustically?

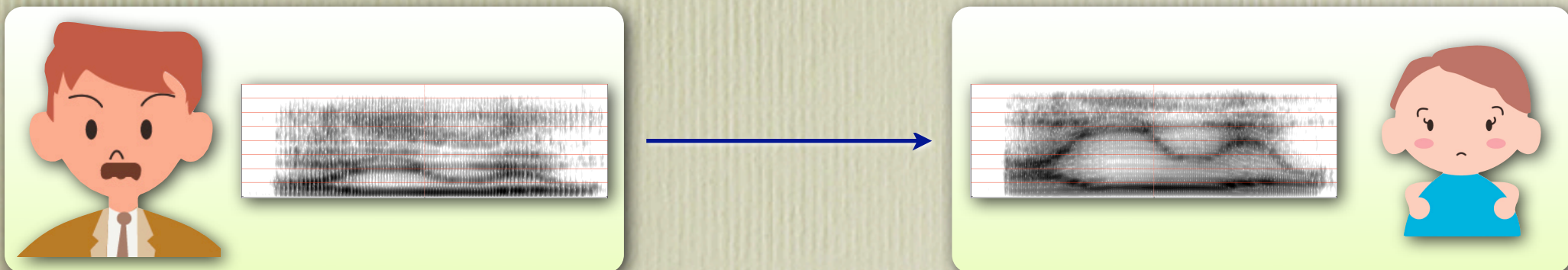
Two answers to the question

Utterance → symbol sequence → utterance



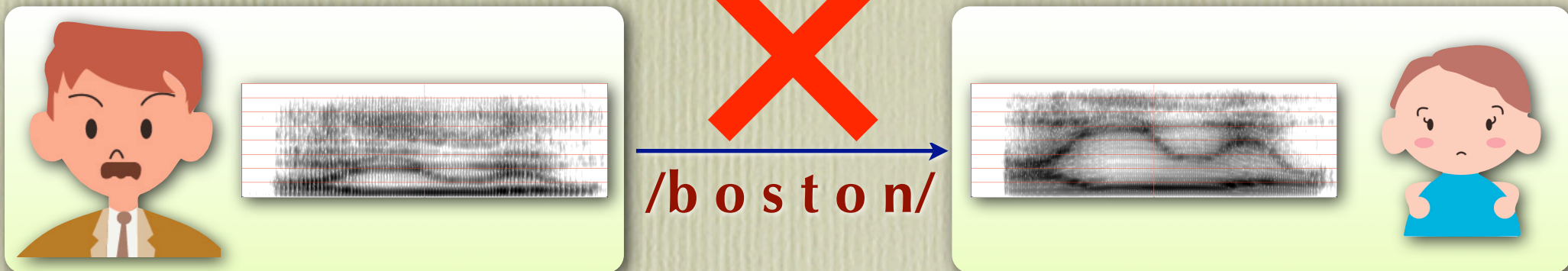
Holistic/relative sound pattern embedded in utterances

- Holistic wordform (Shaywitz'05, Kato'03)
- Word Gestalt (Hayakawa'06)
- Relative pattern (Lieberman'80)



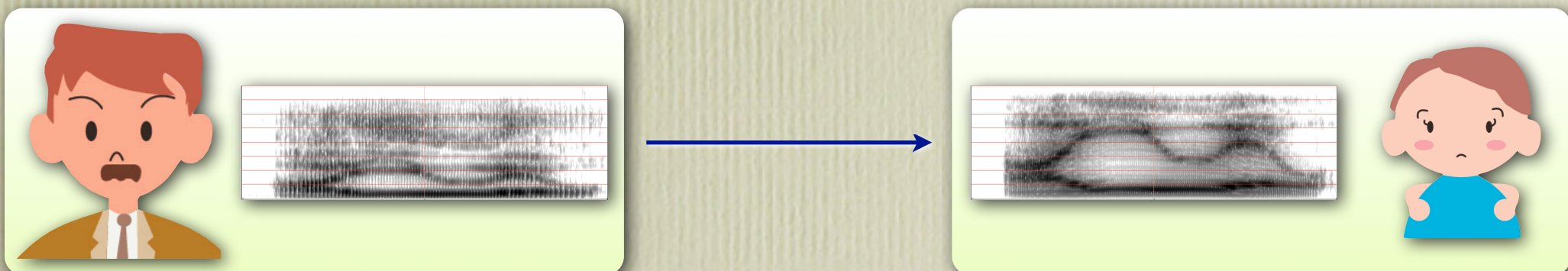
Two answers to the question

Utterance → symbol sequence → utterance



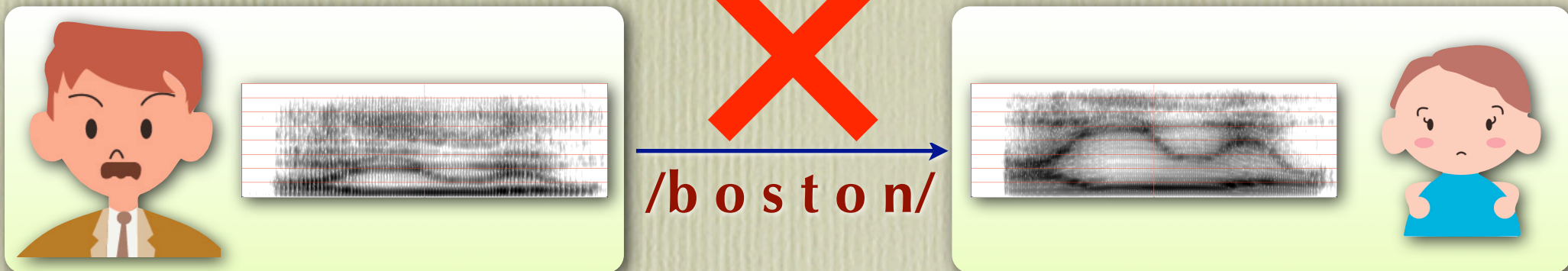
Holistic/relative sound pattern embedded in utterances

- Holistic wordform (Shaywitz'05, Kato'03)
- Word Gestalt (Hayakawa'06)
- Relative pattern (Lieberman'80)



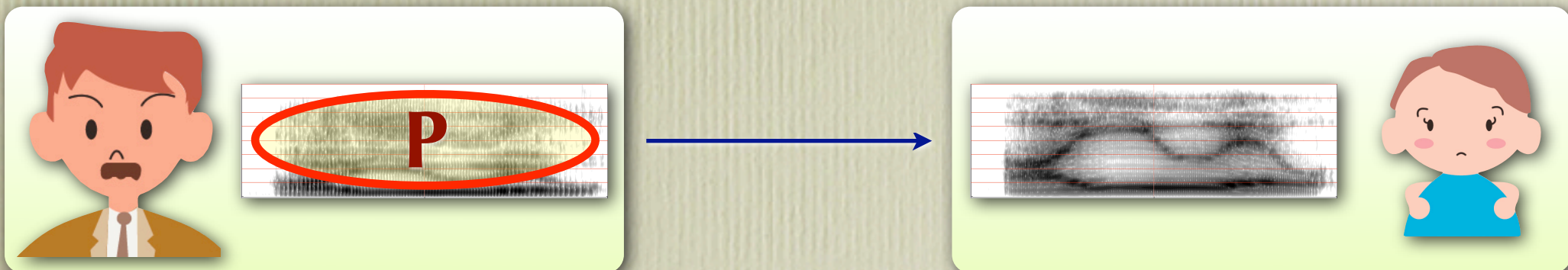
Two answers to the question

Utterance → symbol sequence → utterance



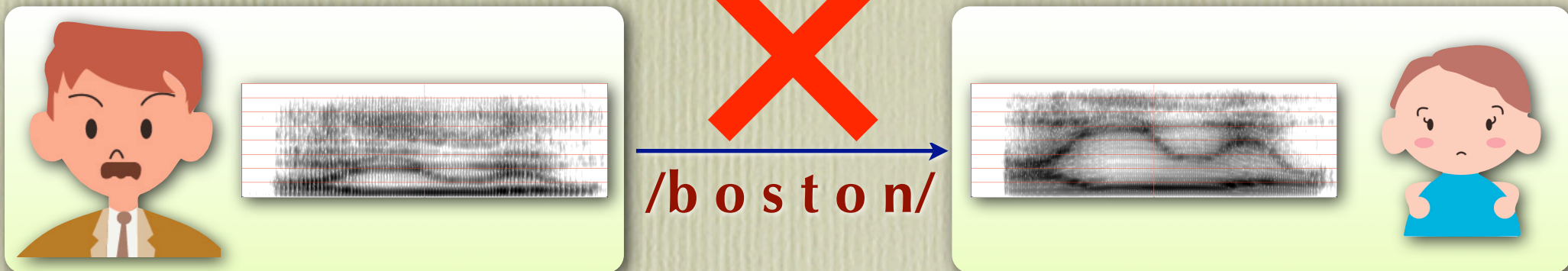
Holistic/relative sound pattern embedded in utterances

- Holistic wordform (Shaywitz'05, Kato'03)
- Word Gestalt (Hayakawa'06)
- Relative pattern (Lieberman'80)



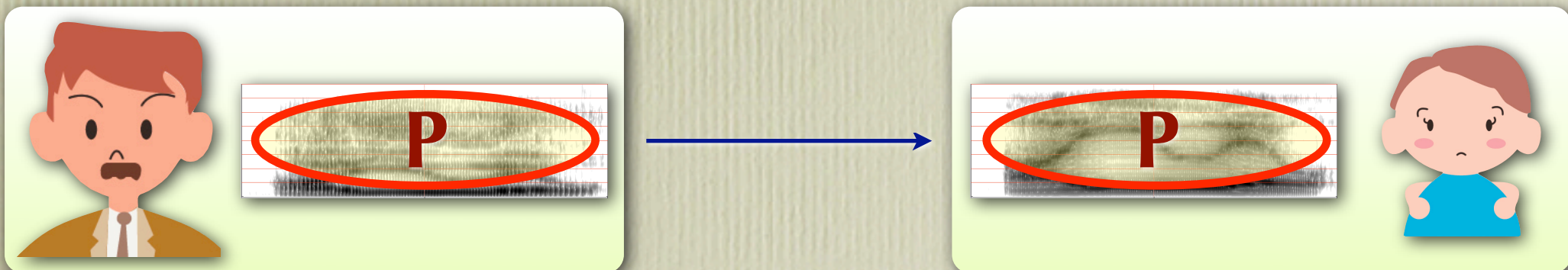
Two answers to the question

Utterance → symbol sequence → utterance



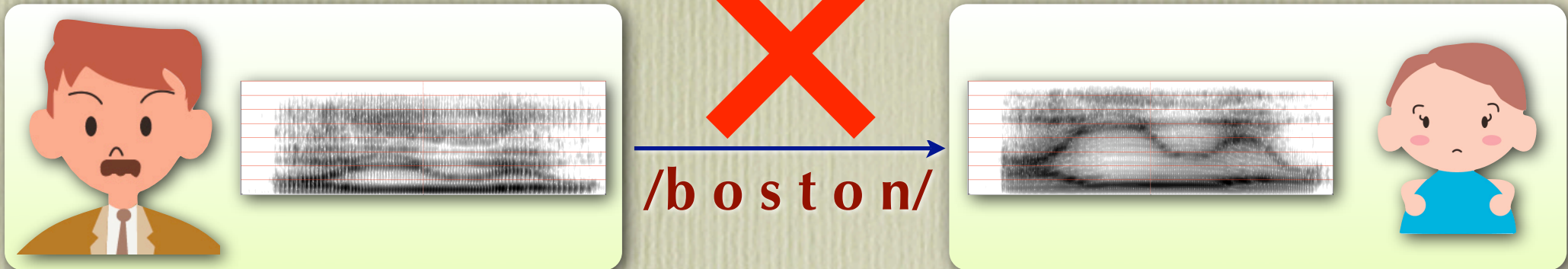
Holistic/relative sound pattern embedded in utterances

- Holistic wordform (Shaywitz'05, Kato'03)
- Word Gestalt (Hayakawa'06)
- Relative pattern (Lieberman'80)



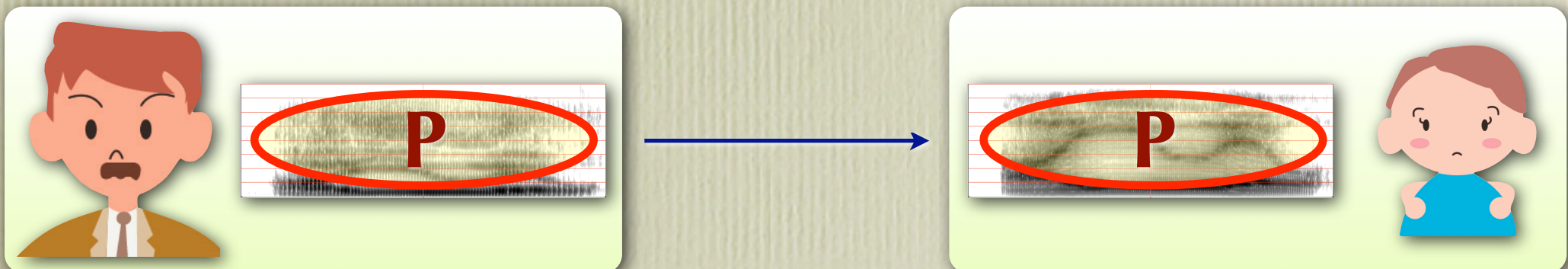
What in voices do infants imitate acoustically?

Utterance → symbol sequence → utterance



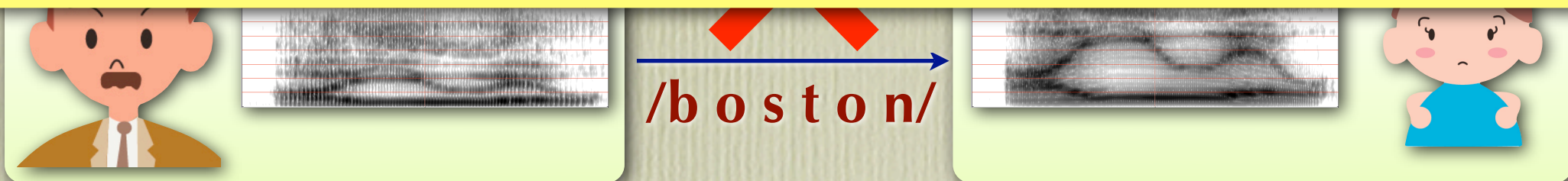
Holistic/relative sound pattern embedded in utterances

- Holistic wordform (Shaywitz'05, Kato'03)
- Word Gestalt (Hayakawa'06)
- Relative pattern (Lieberman'80)



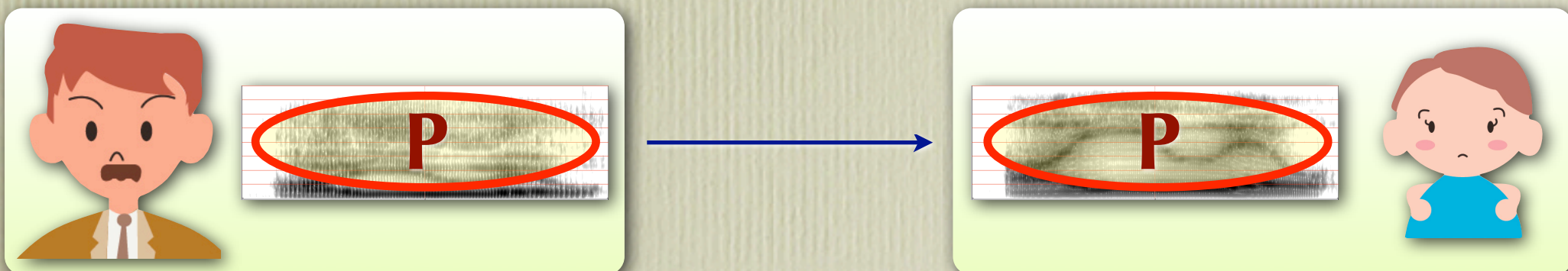
What in voices do infants imitate acoustically?

What is the acoustic definition of word Gestalt?



Holistic/relative sound pattern embedded in utterances

- Holistic wordform (Shaywitz'05, Kato'03)
- Word Gestalt (Hayakawa'06)
- Relative pattern (Lieberman'80)



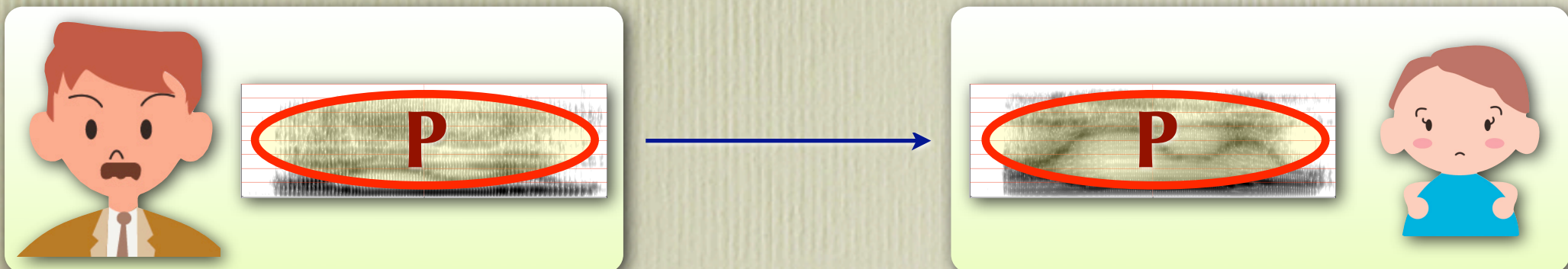
What in voices do infants imitate acoustically?

What is the acoustic definition of word Gestalt?

What is the holistic & speaker-invariant pattern?

 **Holistic/relative sound pattern embedded in utterances**

- Holistic wordform (Shaywitz'05, Kato'03)
- Word Gestalt (Hayakawa'06)
- Relative pattern (Lieberman'80)



Acoustic variability of speech

Speech varies due to physical conditions of vocal organs

Two kinds of variability -- pitch and timbre --

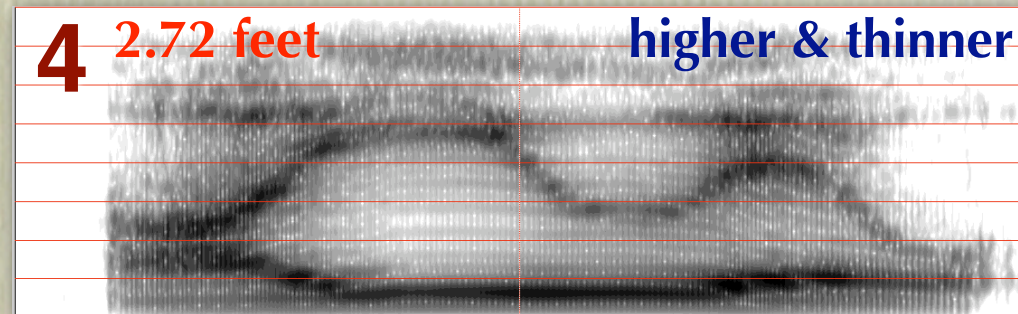
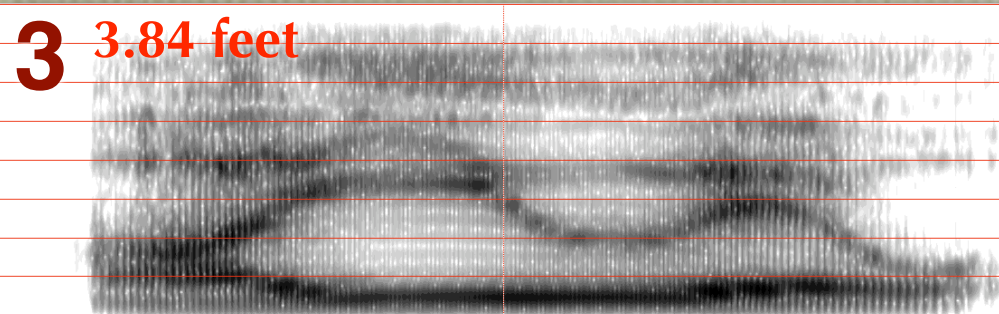
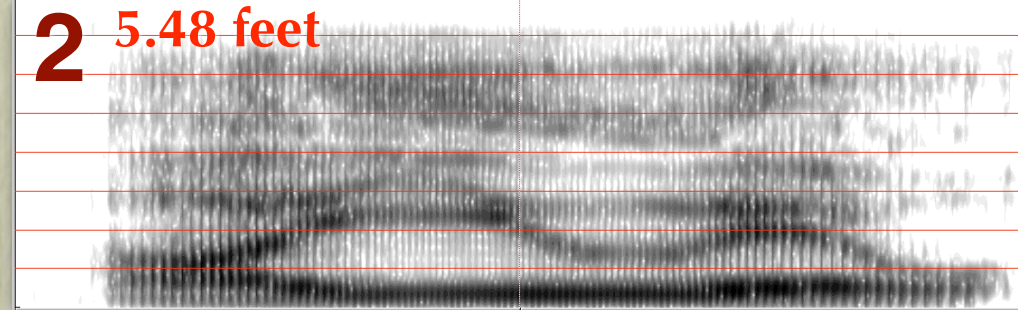
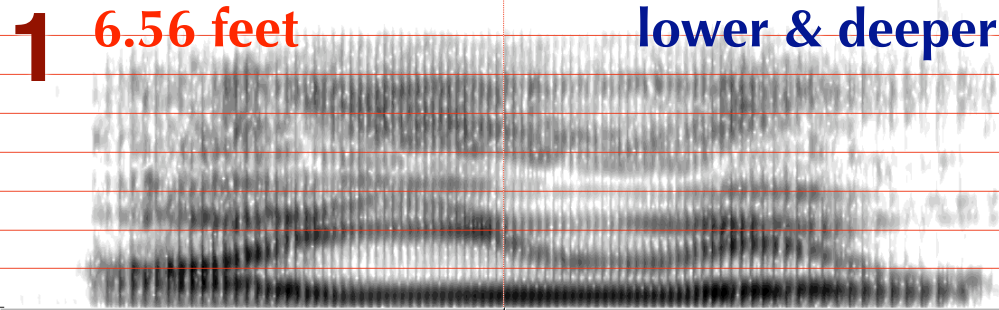
Pitch variability : difference in mass and length of the vocal cords

Timbre variability : difference in size and length of the vocal tract



Four speech samples generated artificially from a single utterance.

Variability in speech acoustics and invariance in speech percepts



Holistic and speaker-invariant pattern

Two kinds of variability -- pitch and timbre --

- Pitch variability : difference in mass and length of the vocal cords.
- Timbre variability : difference in size and length of the vocal tracts

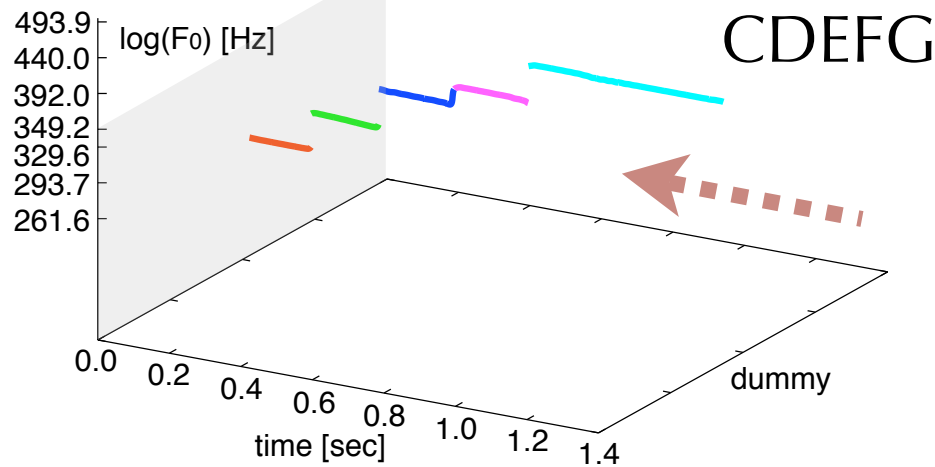
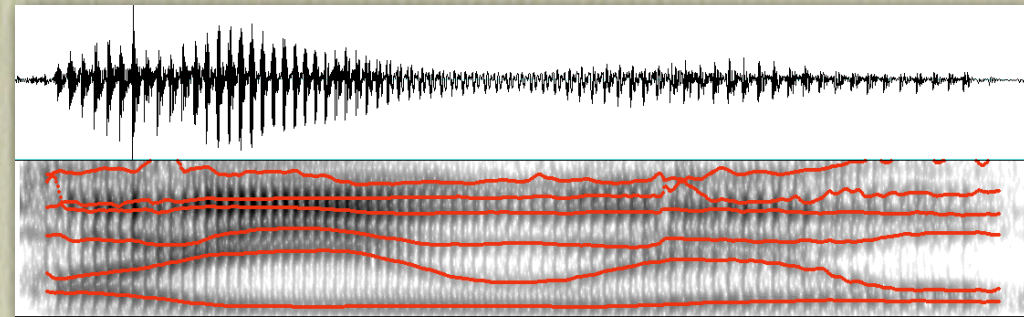
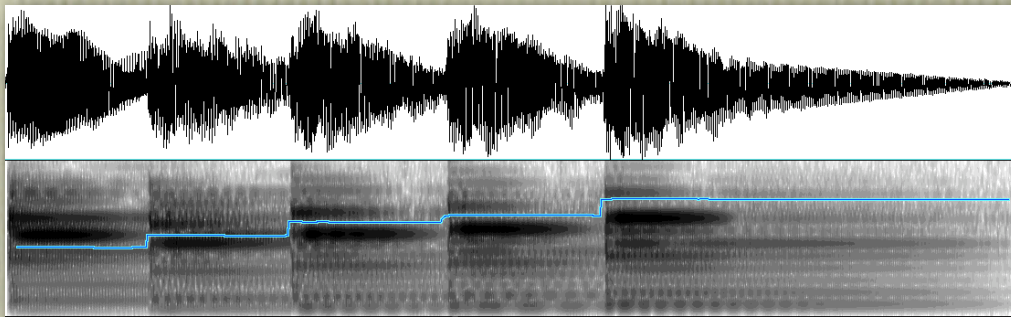


A very simple answer for pitch variability

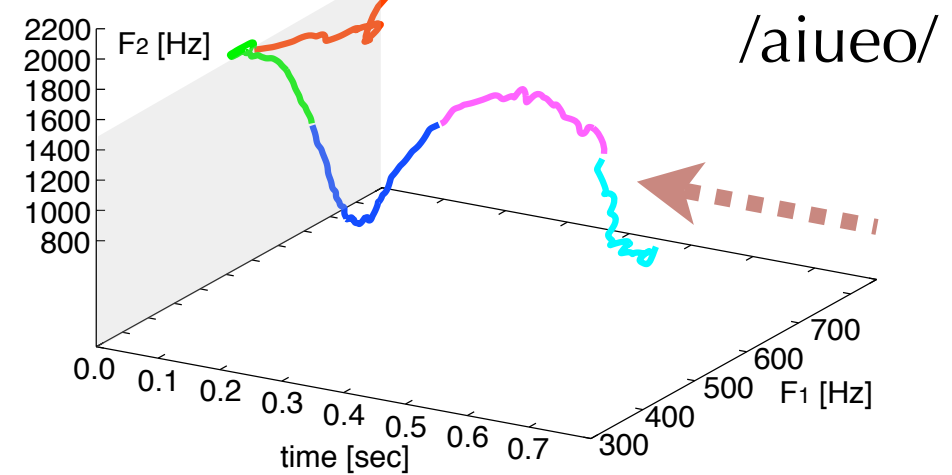
- Relative and dynamic pitch patterns are invariant.
 - Average pitch is different among speakers.
 - But its dynamic pattern is the same among speakers.
- Relative pitch patterns are invariant wrt. a change of key.



Dynamic patterns in pitch and timbre



CDEFG

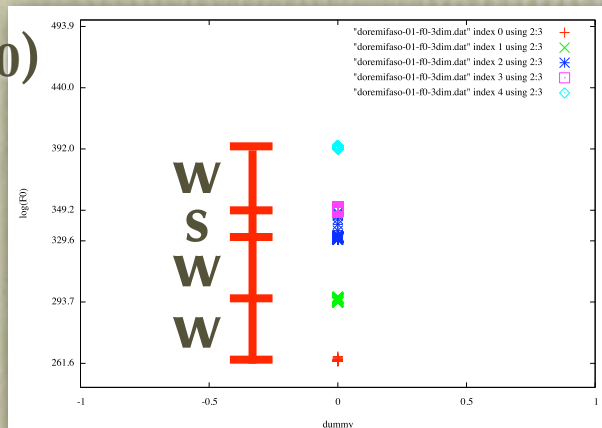


/aiueo/

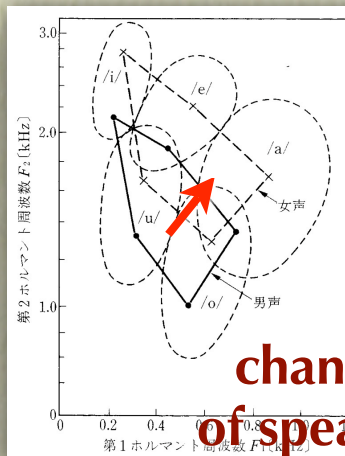
pitch-based melody

timbre-based melody

$\log(F_0)$

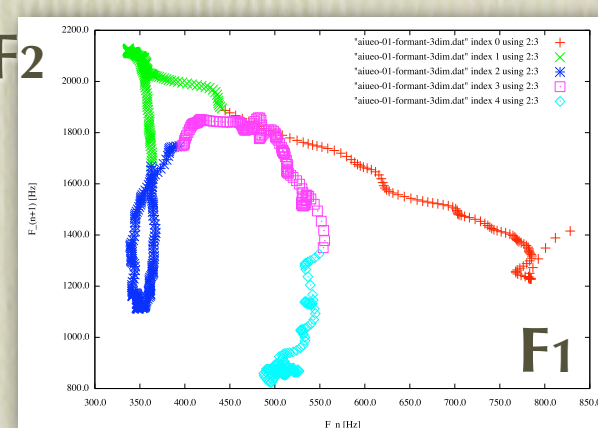


change of key



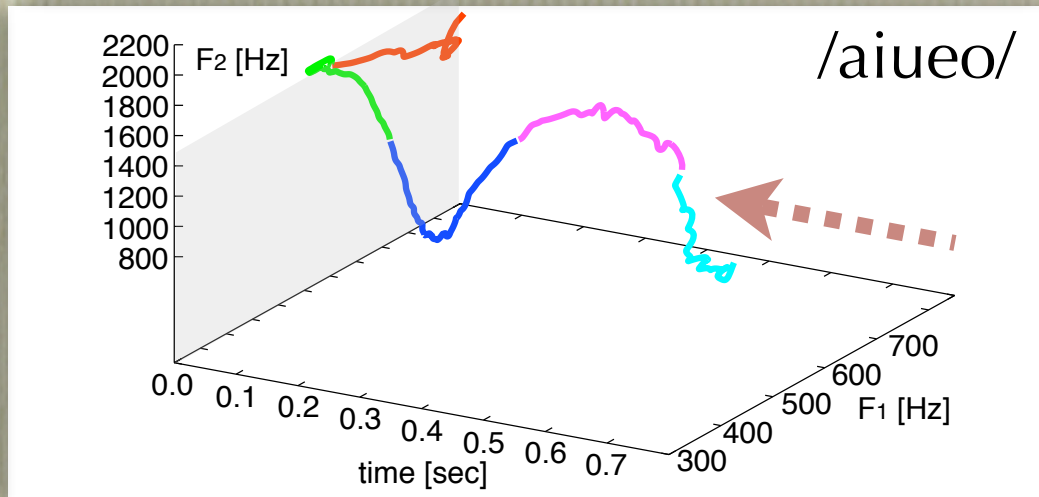
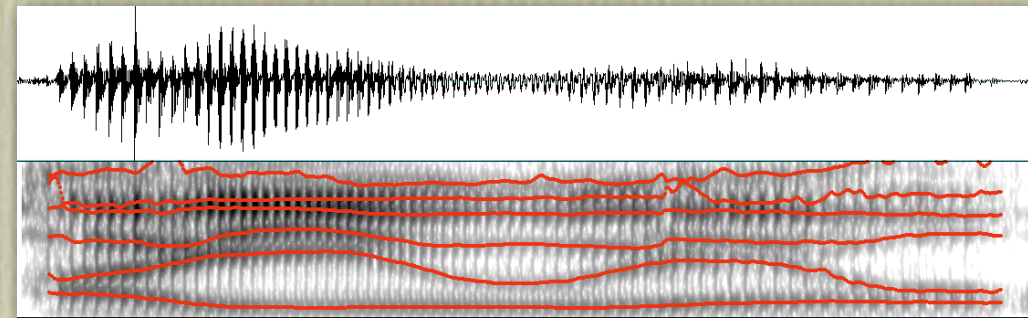
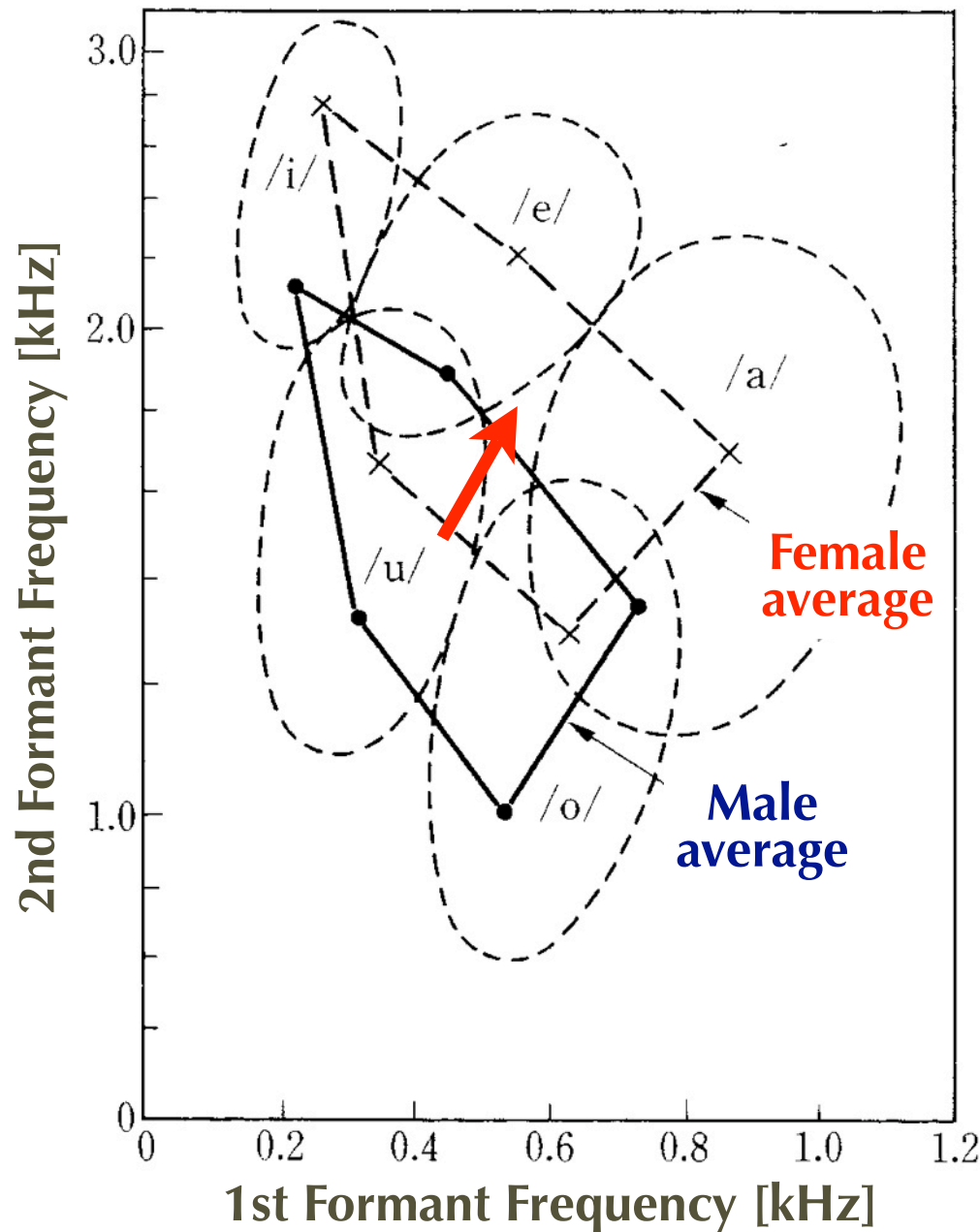
change of speaker

F_2

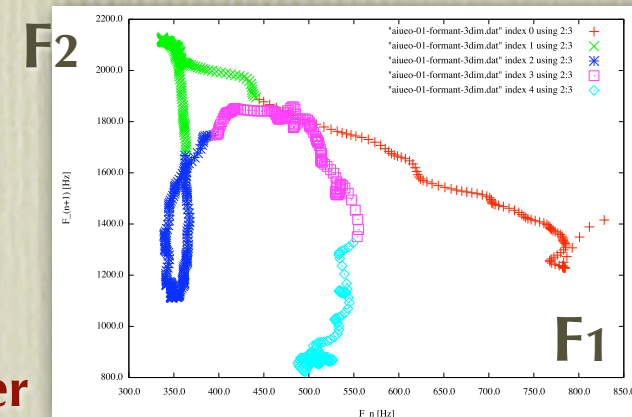
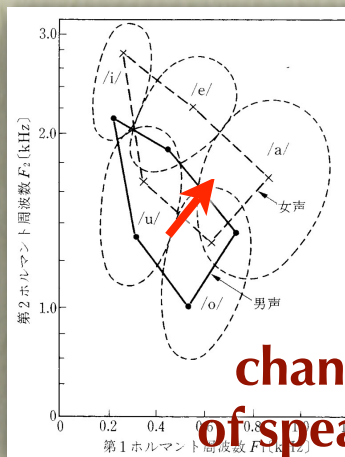


F_1

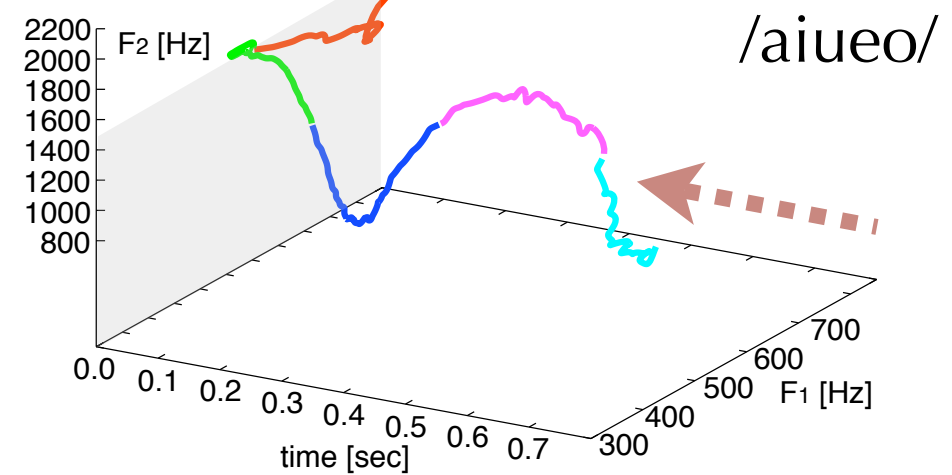
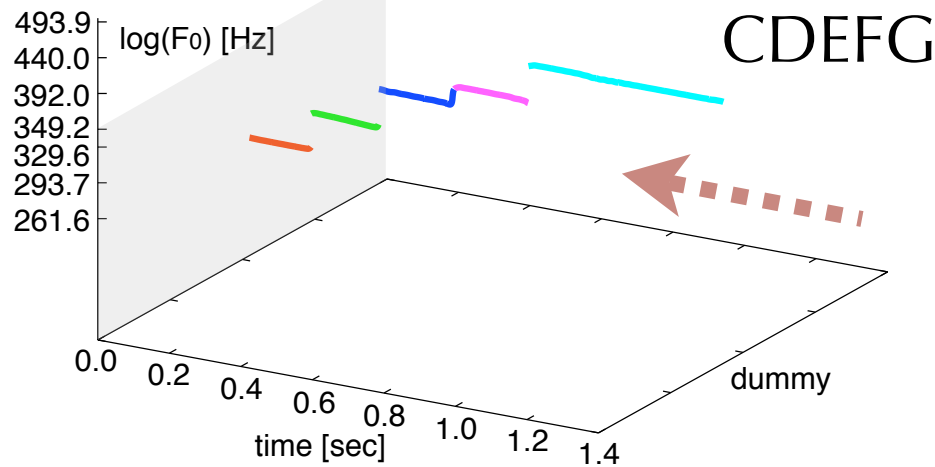
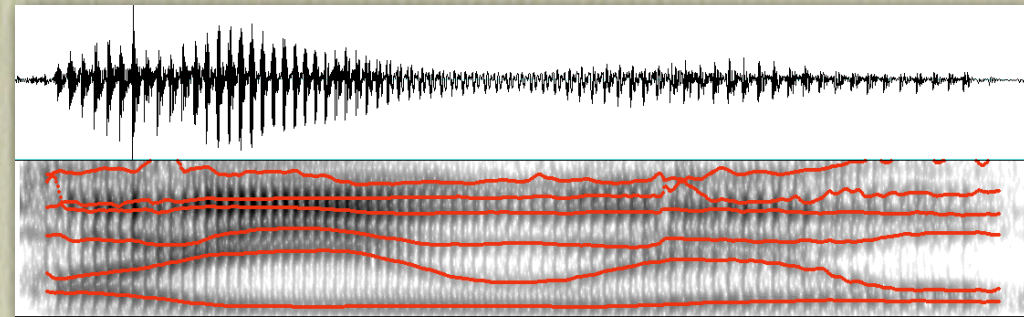
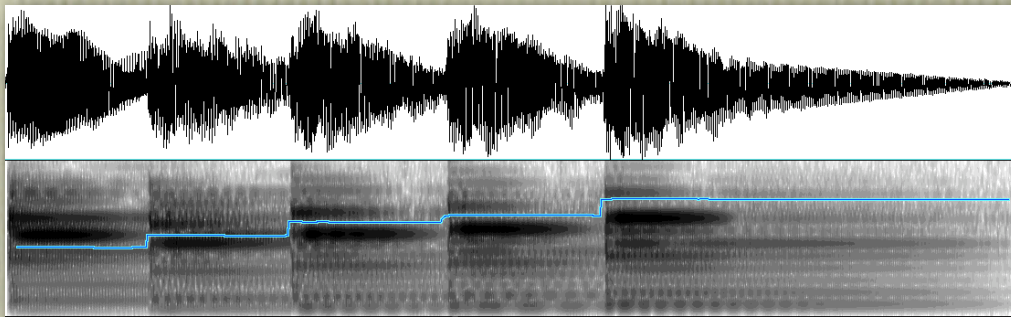
Dynamic patterns in pitch and timbre



timbre-based melody



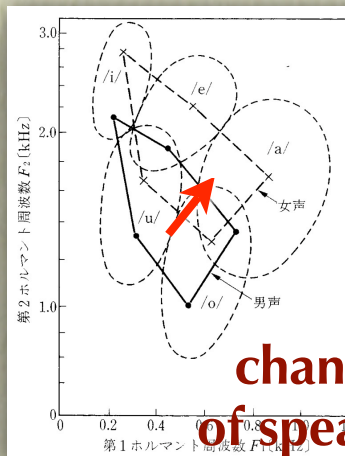
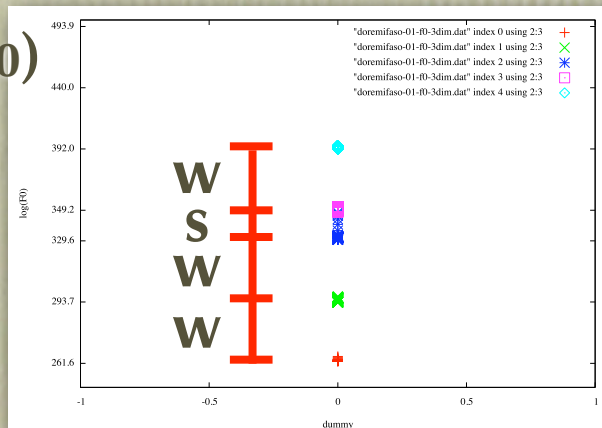
Dynamic patterns in pitch and timbre



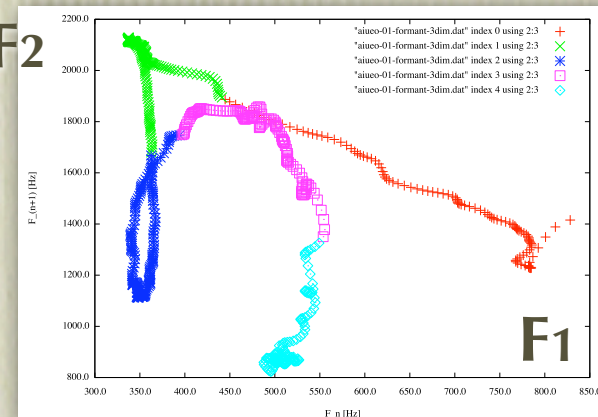
pitch-based melody

timbre-based melody

log(F0)



F2



F1

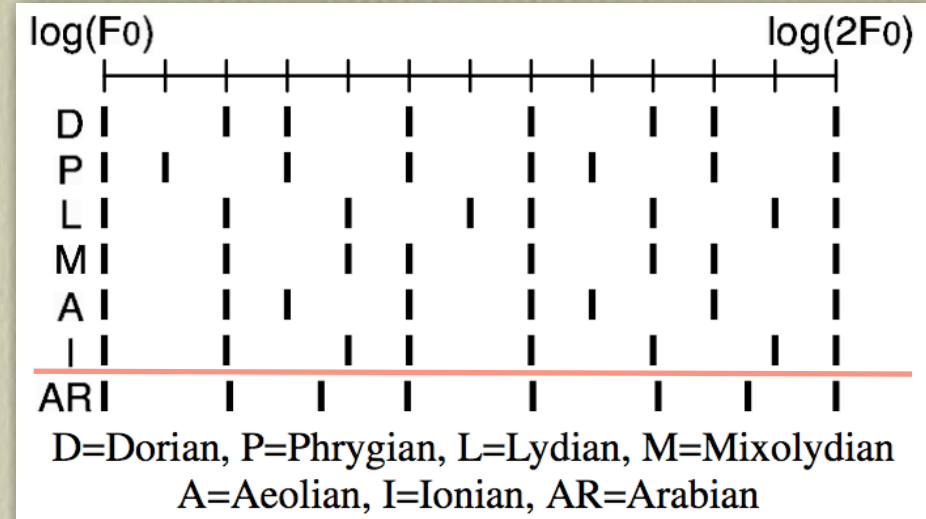
change of key

change of speaker

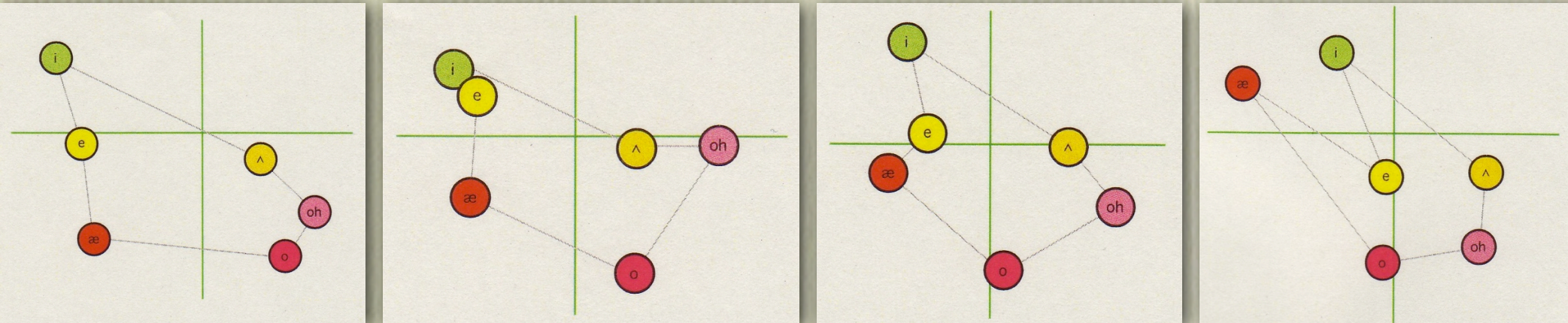
A good similarity between music and speech

Variation of sound arrangement in pitch

- Classical church music
 - Dorian, Phrygian, Lydian,
- Major and Minor
 - Ionian = Major, Aeolian = Minor
- Arabic



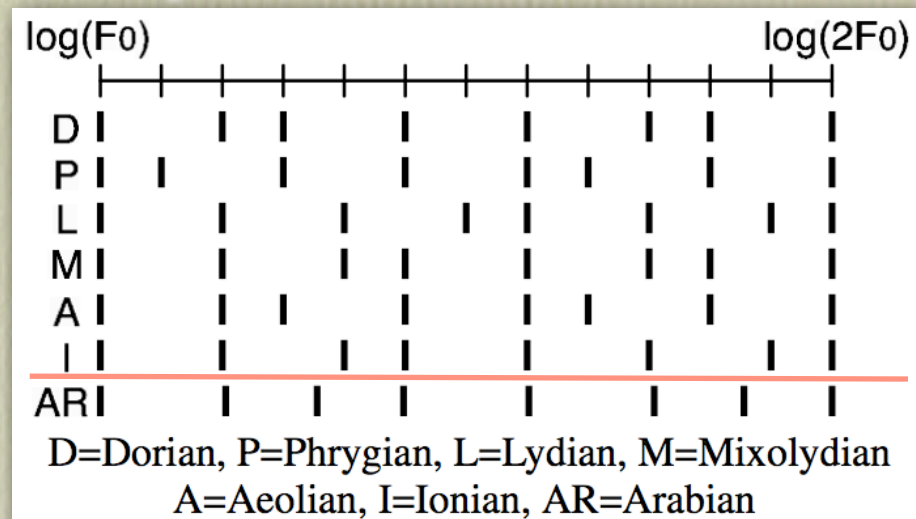
Variation of sound arrangement in timbre



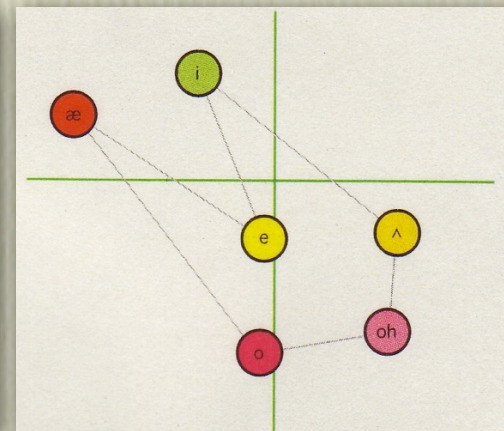
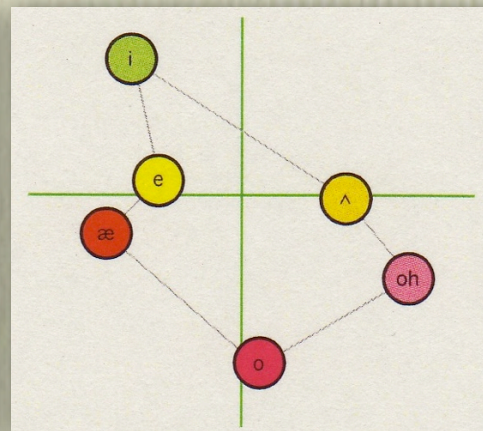
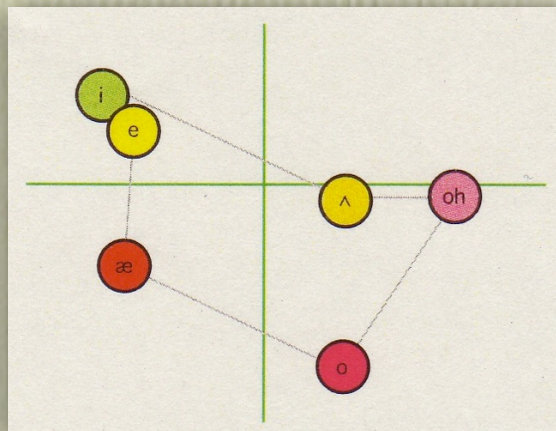
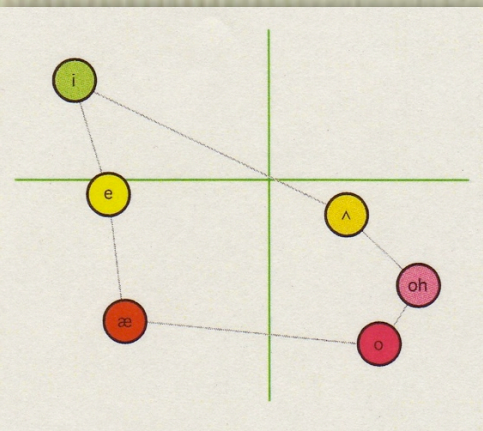
A good similarity between music and speech

Variation of sound arrangement in pitch

- Classical church music
 - Dorian, Phrygian, Lydian,
- Major and Minor
 - Ionian = Major, Aeolian = Minor
- Arabic



Variation of sound arrangement in timbre



Williamsport, PA

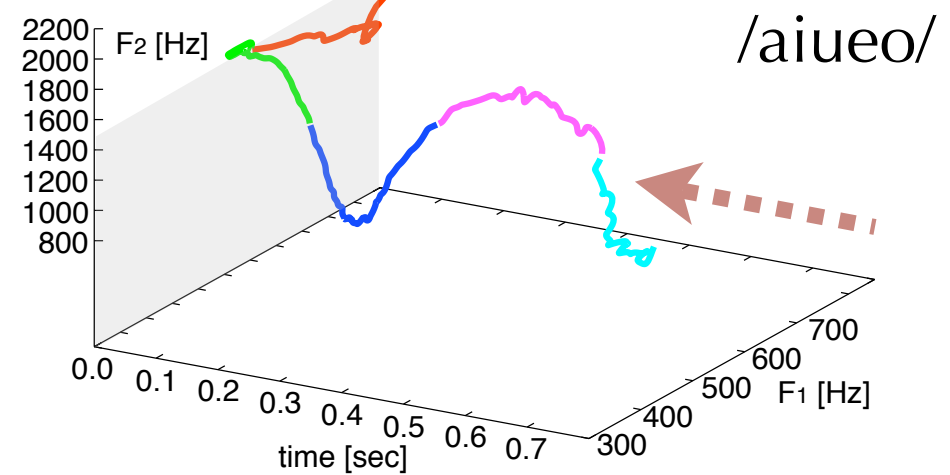
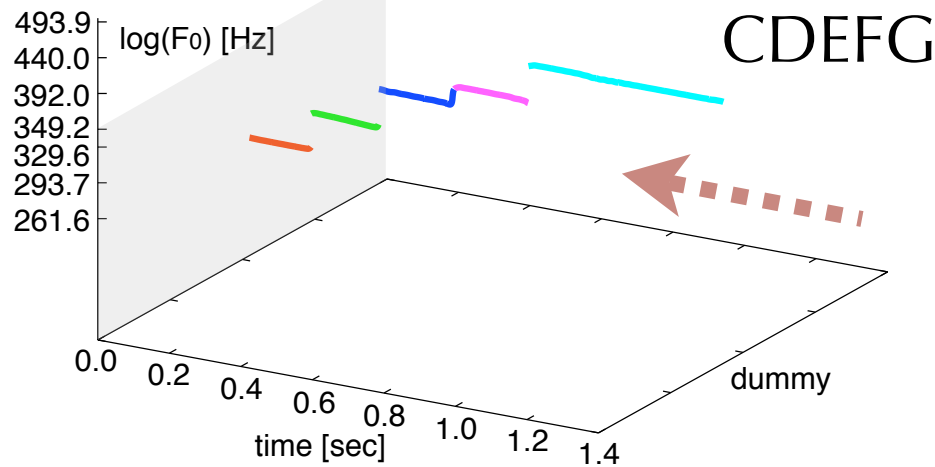
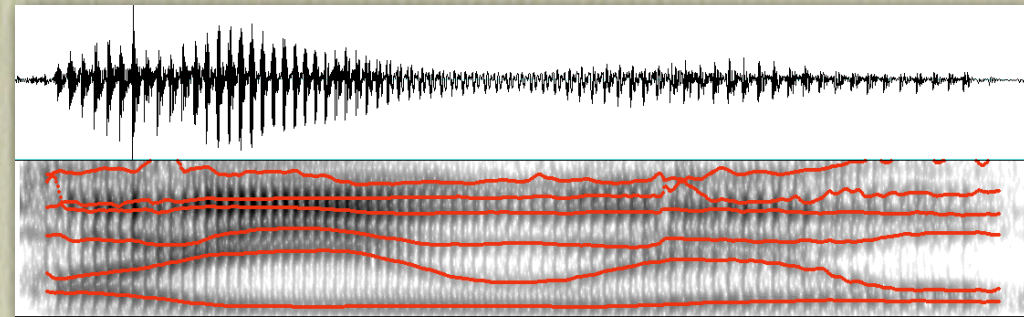
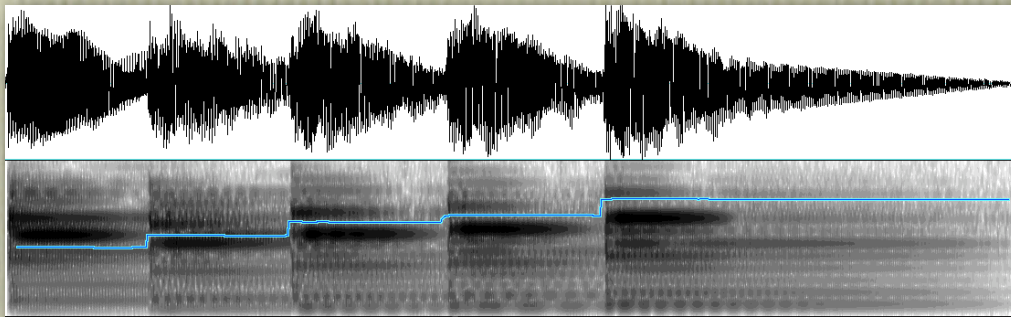
Chicago, IL

Ann Arbor, MI

Rochester, NY

From *Atlas of North American English* (W. Labov et al.)

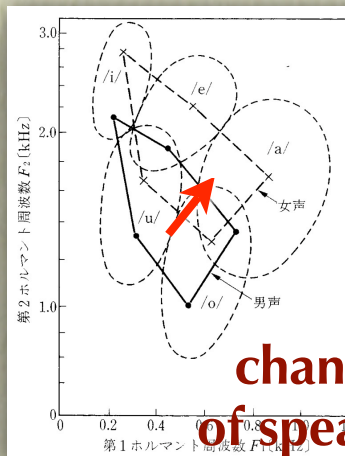
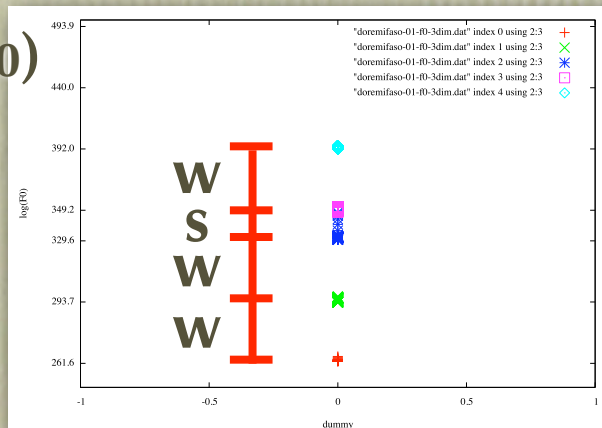
Dynamic patterns in pitch and timbre



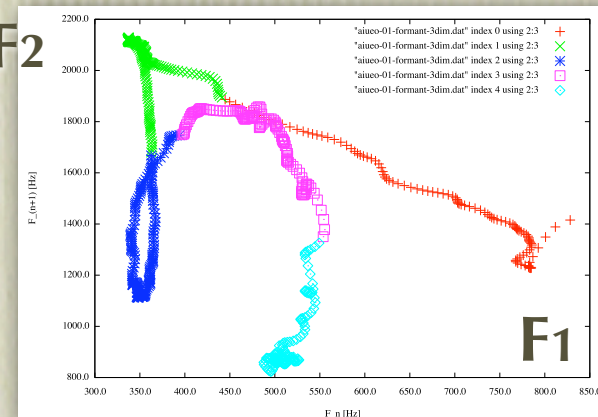
pitch-based melody

timbre-based melody

log(F0)



F2



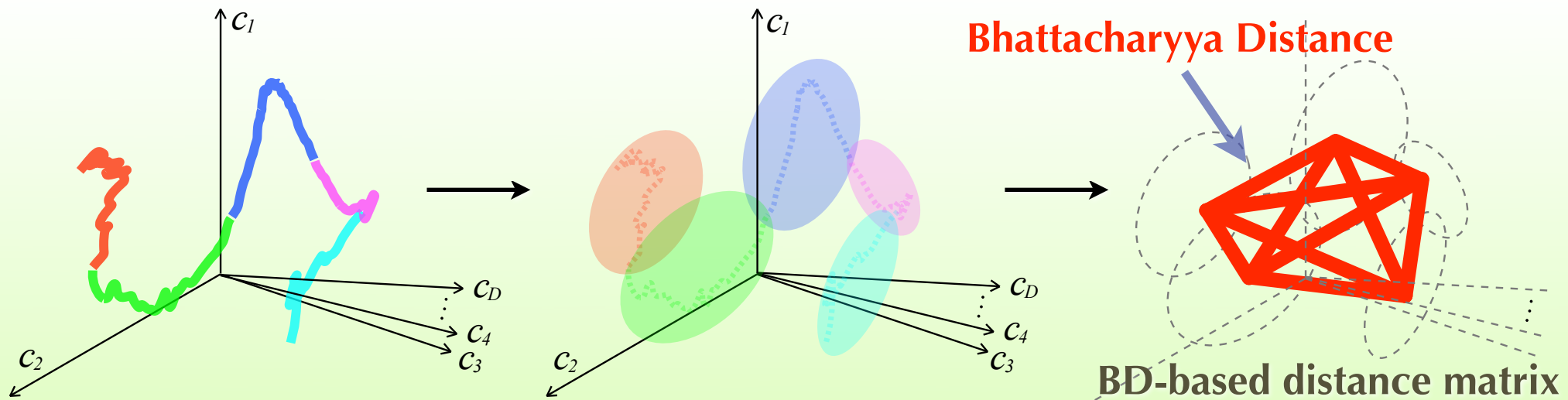
F1

change of key

change of speaker

Speaker-invariant pattern in timbre

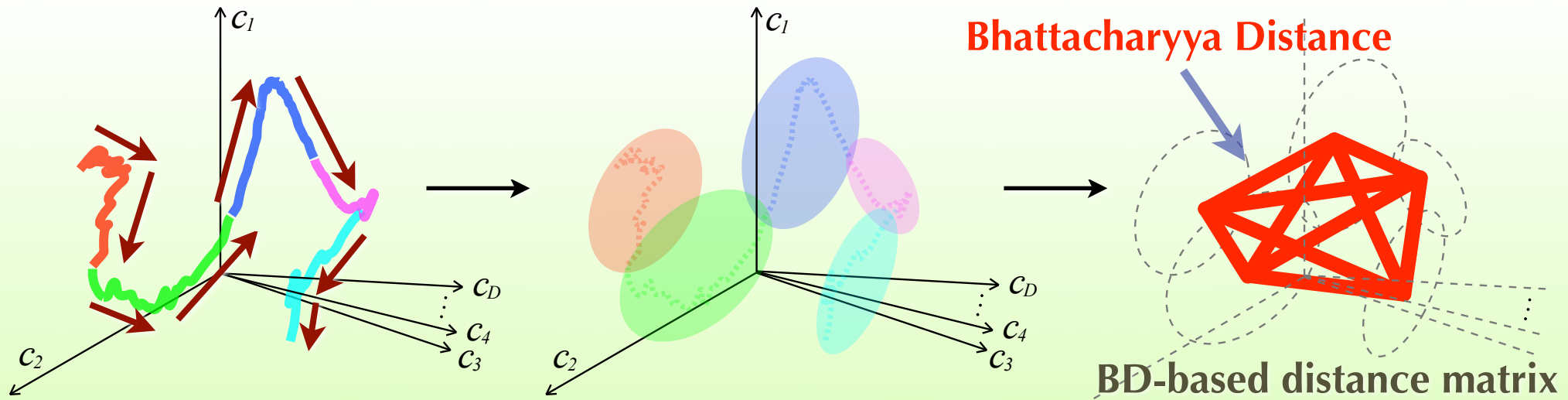
Timbre sequence $\rightarrow$ distribution seq. $\rightarrow$ distance matrix



- Timbre trajectory is rotated due to vocal tract length difference.
- Directional features are strongly speaker-dependent.
- Only contrasts are extracted from an utterance.
- Distribution-to-distribution distances are calculated.
- Absolute properties of timbre are completely discarded.
- A full set of contrasts/distances can form a geometrical structure.
- This geometrical structure is holistic and speaker-invariant.

Speaker-invariant pattern in timbre

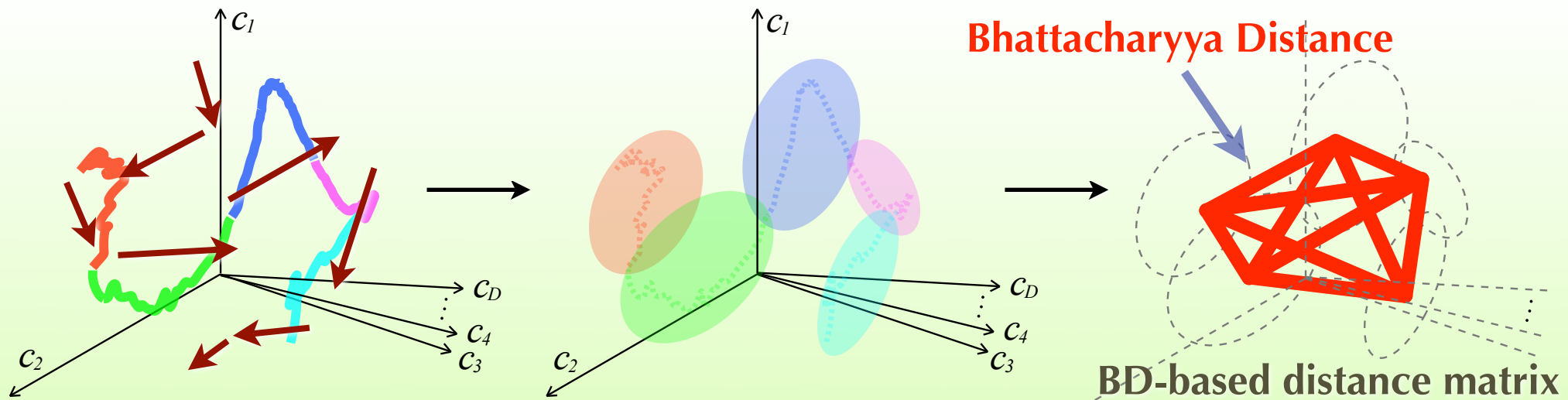
Timbre sequence $\rightarrow$ distribution seq. $\rightarrow$ distance matrix



- Timbre trajectory is rotated due to vocal tract length difference.
- Directional features are strongly speaker-dependent.
- Only contrasts are extracted from an utterance.
- Distribution-to-distribution distances are calculated.
- Absolute properties of timbre are completely discarded.
- A full set of contrasts/distances can form a geometrical structure.
- This geometrical structure is holistic and speaker-invariant.

Speaker-invariant pattern in timbre

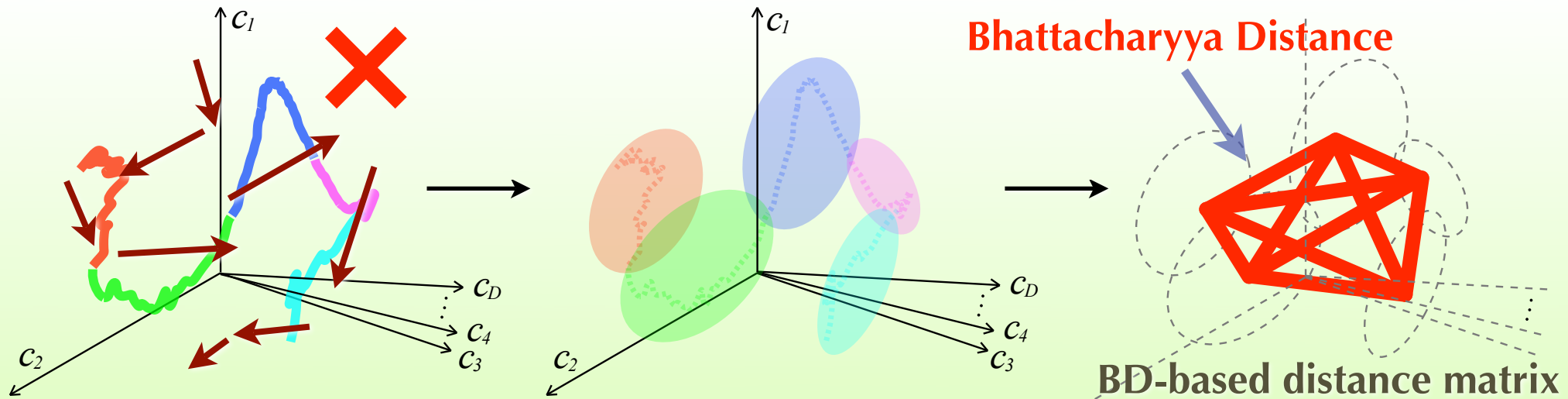
Timbre sequence $\rightarrow$ distribution seq. $\rightarrow$ distance matrix



- Timbre trajectory is rotated due to vocal tract length difference.
- Directional features are strongly speaker-dependent.
- Only contrasts are extracted from an utterance.
- Distribution-to-distribution distances are calculated.
- Absolute properties of timbre are completely discarded.
- A full set of contrasts/distances can form a geometrical structure.
- This geometrical structure is holistic and speaker-invariant.

Speaker-invariant pattern in timbre

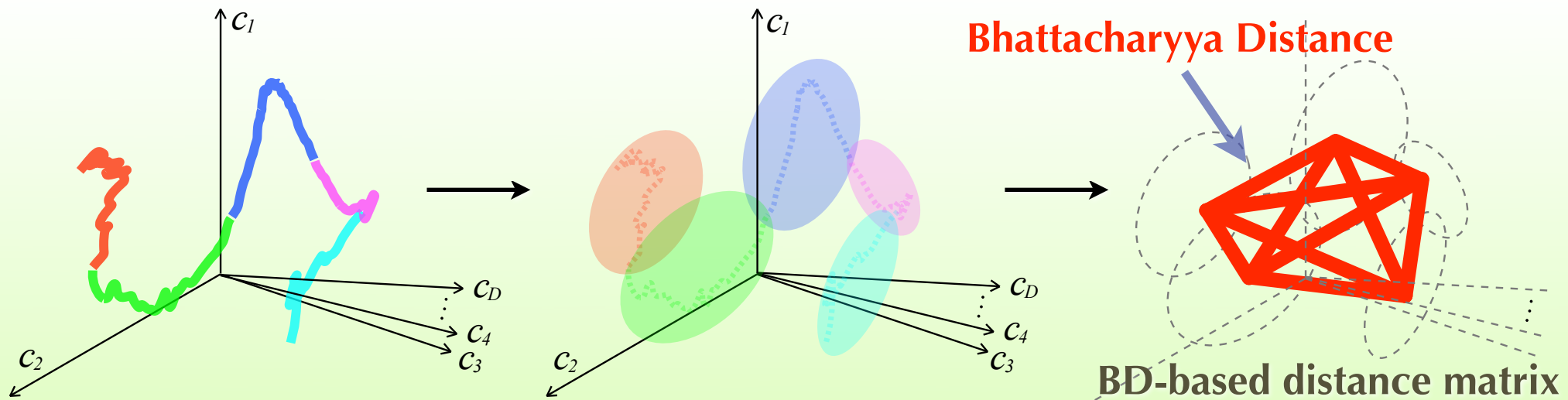
Timbre sequence $\rightarrow$ distribution seq. $\rightarrow$ distance matrix



- Timbre trajectory is rotated due to vocal tract length difference.
- Directional features are strongly speaker-dependent.
- Only contrasts are extracted from an utterance.
- Distribution-to-distribution distances are calculated.
- Absolute properties of timbre are completely discarded.
- A full set of contrasts/distances can form a geometrical structure.
- This geometrical structure is holistic and speaker-invariant.

Speaker-invariant pattern in timbre

Timbre sequence $\rightarrow$ distribution seq. $\rightarrow$ distance matrix

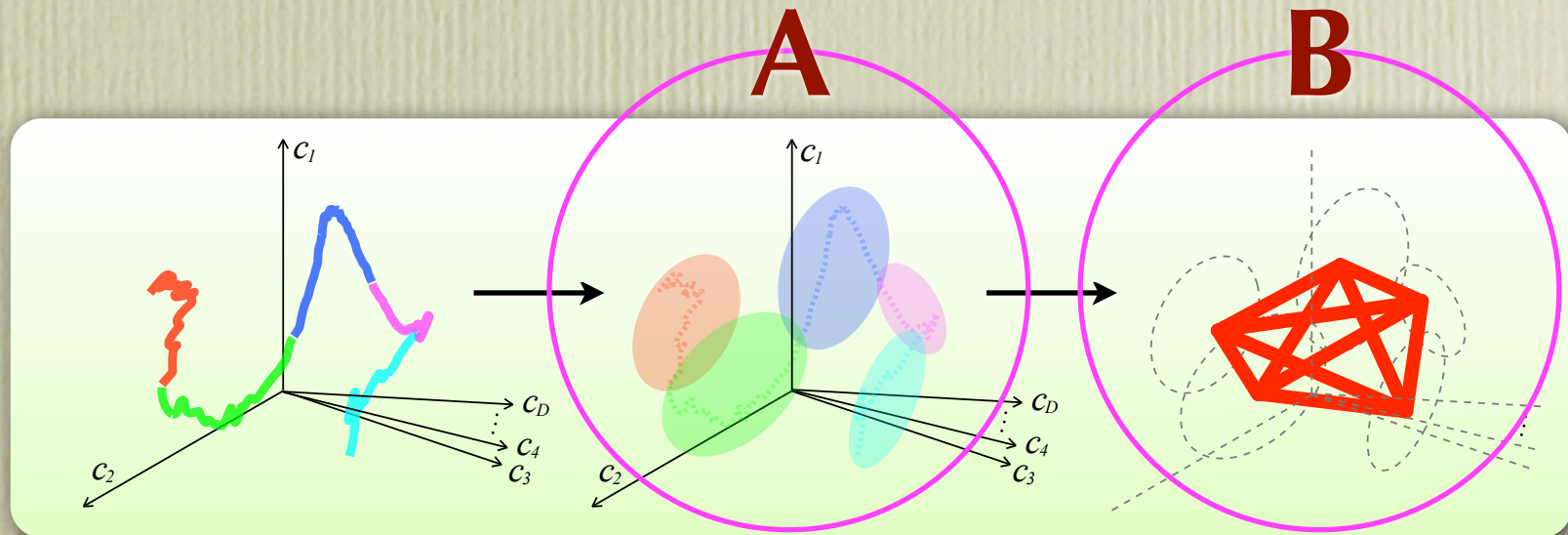


- Timbre trajectory is rotated due to vocal tract length difference.
- Directional features are strongly speaker-dependent.
- Only contrasts are extracted from an utterance.
- Distribution-to-distribution distances are calculated.
- Absolute properties of timbre are completely discarded.**
- A full set of contrasts/distances can form a geometrical structure.
- This geometrical structure is holistic and speaker-invariant.

An isolated word recognition experiment

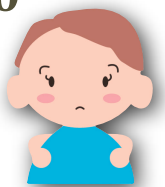
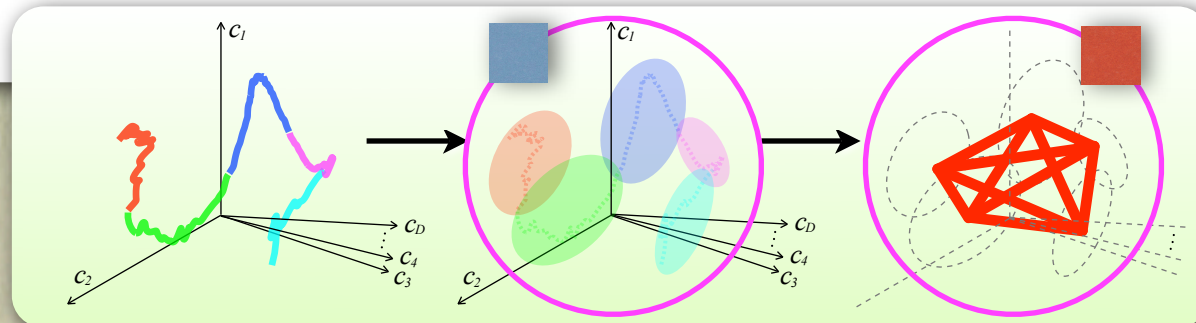
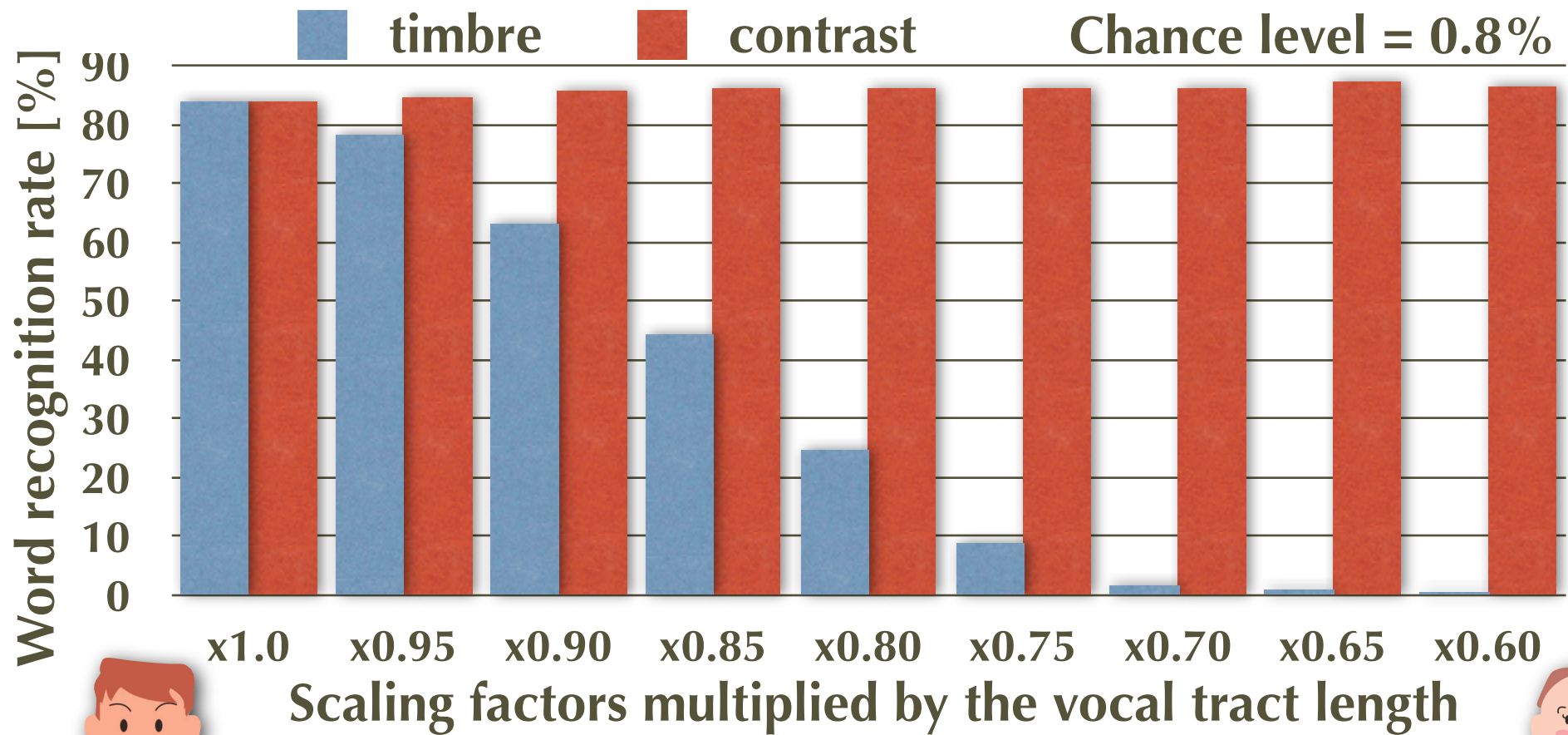
Experimental conditions

- Word = V_1 - V_2 - V_3 - V_4 - V_5 V_i = one of the 5 Japanese vowels
 - /aiueo/, /eoau/, /oeau/, /iouai/, etc Vocab. size = 120 (CL = 0.8%)
- Two kinds of 120 word templates are stored in speech recognizers
 - A:** timbre-based word templates made with adult speakers' data
 - Timbre can be changed easily by size and length of the vocal tract.
 - B:** timbre-contrast-based word templates made with adults' data
 - Timbre contrasts are invariant wrt. size and length of the vocal tract.



An isolated word recognition experiment

Timbre templates vs. timbre-contrast templates



Conclusion

What if the parents of **identical twins** divorce?

- Each of the parents takes in a twin.
- What kind of pronunciation will the twins acquire 10 years later?



Conclusion

What if the parents of **identical twins** divorce?

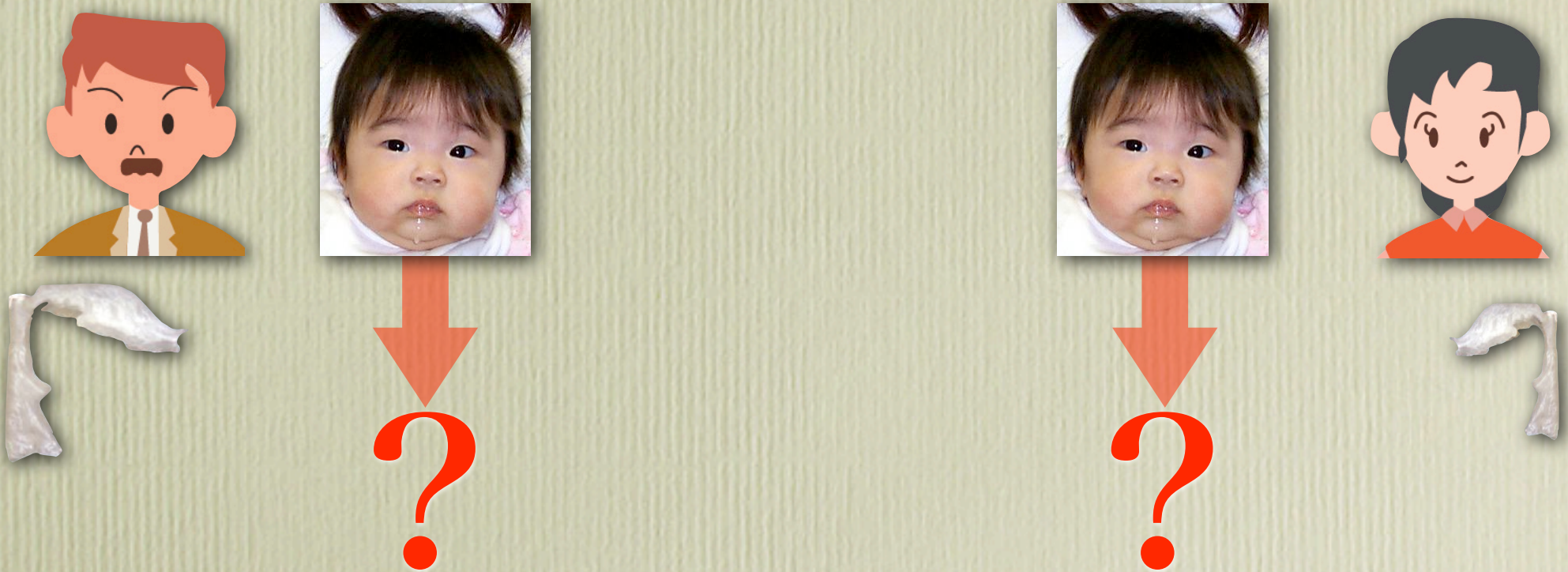
- Each of the parents takes in a twin.
- What kind of pronunciation will the twins acquire 10 years later?



Conclusion

What if the parents of **identical twins** divorce?

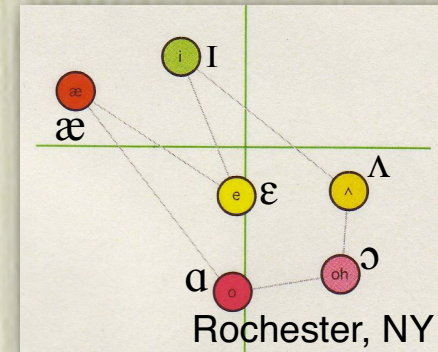
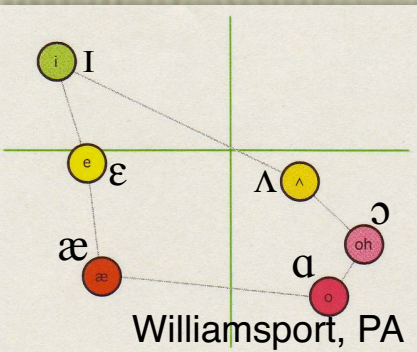
- Each of the parents takes in a twin.
- What kind of pronunciation will the twins acquire 10 years later?



Conclusion

What if the parents of **identical twins** divorce?

- Each of the parents takes in a twin.
- What kind of pronunciation will the twins acquire 10 years later?



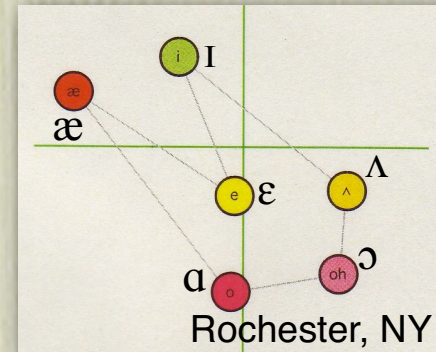
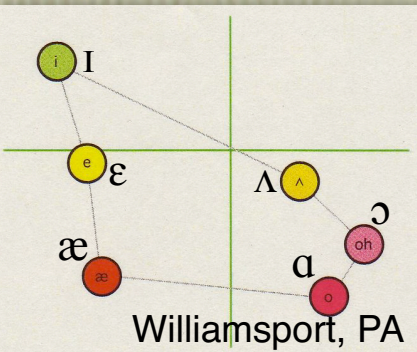
Conclusion

What if the parents of **identical twins** divorce?

- Each of the parents takes in a twin.
- What kind of pronunciation will the twins acquire 10 years later?



Gender difference is a timbre difference.



Conclusion

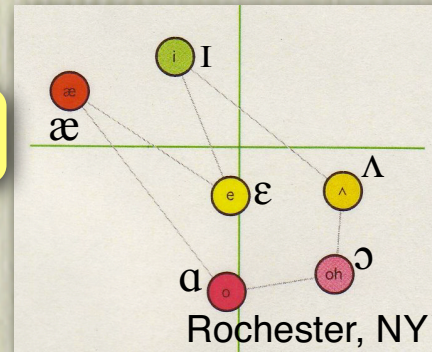
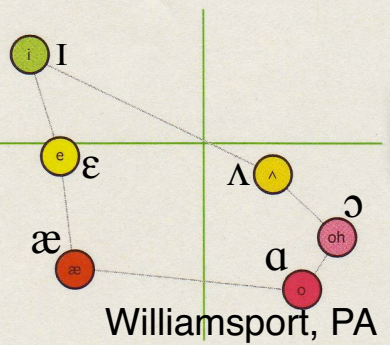
What if the parents of **identical twins** divorce?

- Each of the parents takes in a twin.
- What kind of pronunciation will the twins acquire 10 years later?



Gender difference is a timbre difference.

Regional difference is a timbre difference.



Conclusion

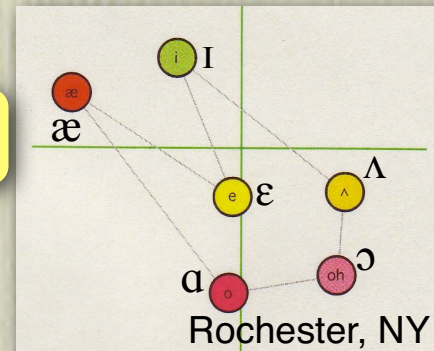
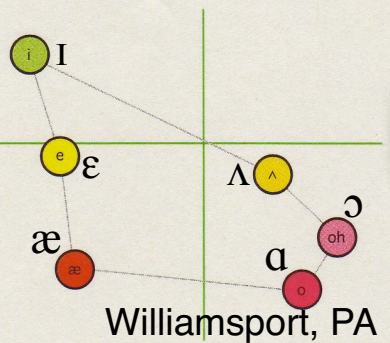
What if the parents of **identical twins** divorce?

- Each of the parents takes in a twin.
- What kind of pronunciation will the twins acquire 10 years later?



Gender difference is a timbre difference.

Regional difference is a timbre difference.



Conclusion

What if the parents of identical twins divorce?

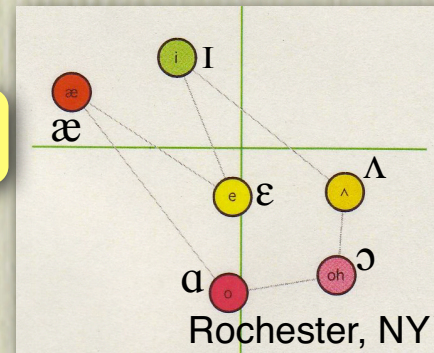
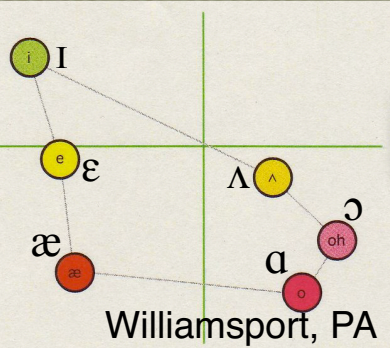
- Each of the parents takes in a twin.
- What kind of pronunciation will the twins acquire 10 years later?

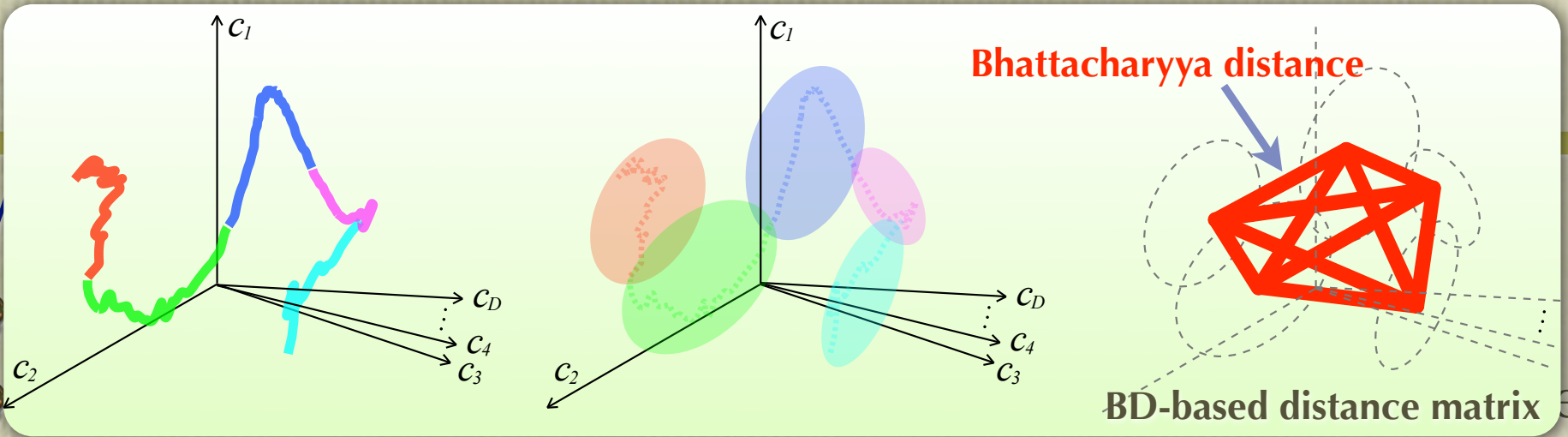


Gender difference is a timbre difference.

Regional difference is a timbre difference.

Myna birds imitate the sounds.
Infants imitate the sound system.

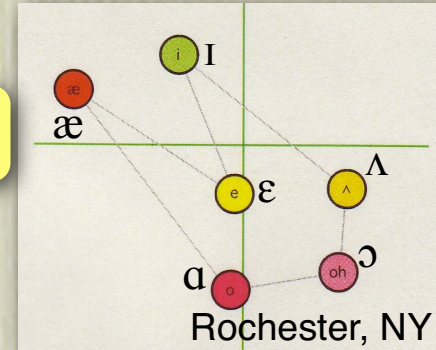
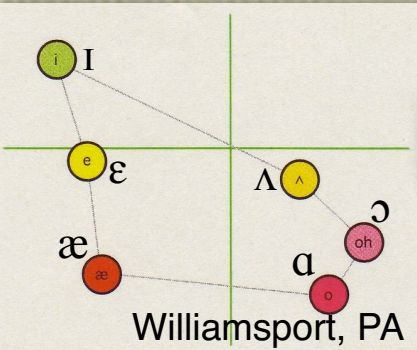


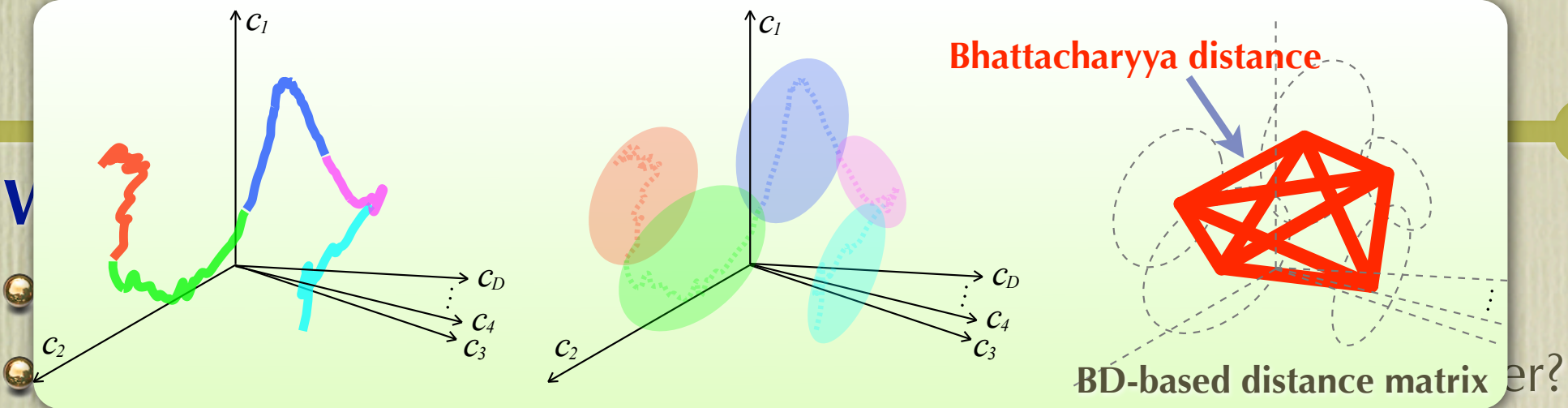


Gender difference is a timbre difference.

Regional difference is a timbre difference.

**Myna birds imitate the sounds.
Infants imitate the sound system.**



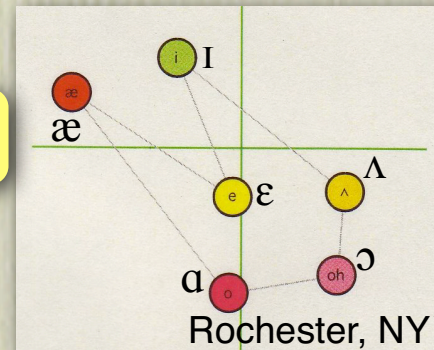
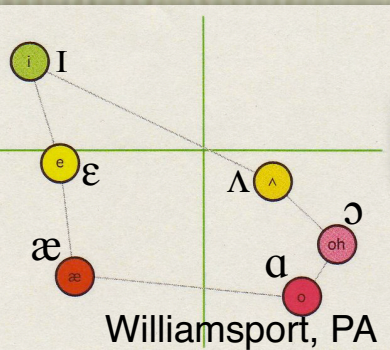


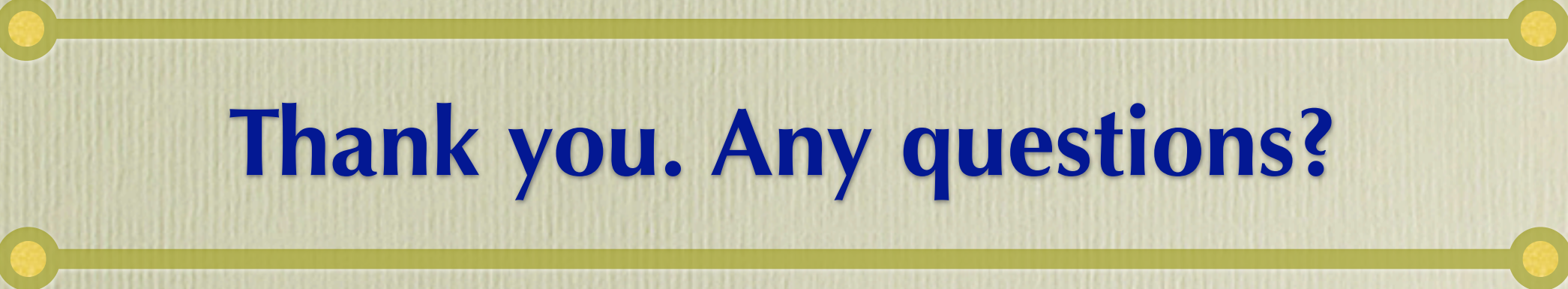
What in voices do infants imitate acoustically?

Gender difference is a timbre difference.

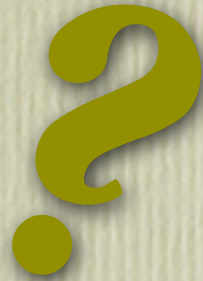
Regional difference is a timbre difference.

Myna birds imitate the sounds.
Infants imitate the sound system.





Thank you. Any questions?



Thank you. Any questions?

