Modeling vocal tract evolution

Bart de Boer
Vrije Universiteit Brussel
Universiteit van Amsterdam

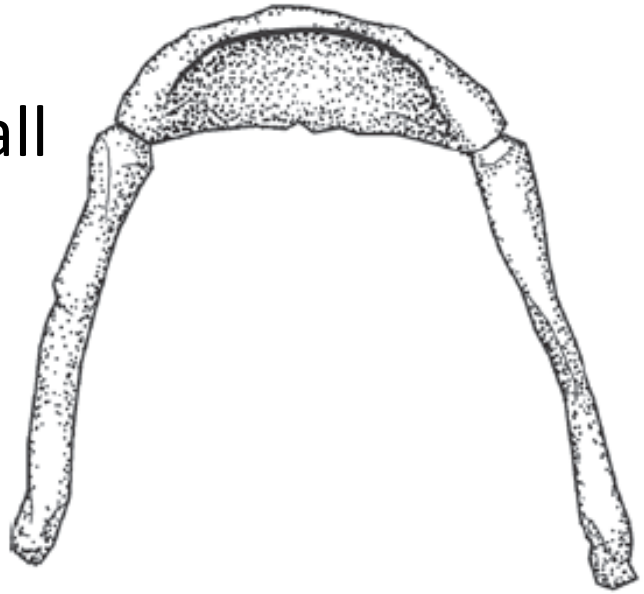
In this talk

- An overview of the evolution of the vocal tract
- Illustrations of the use of computer models
- A brief overview of what has been achieved
- And an interesting open problem…

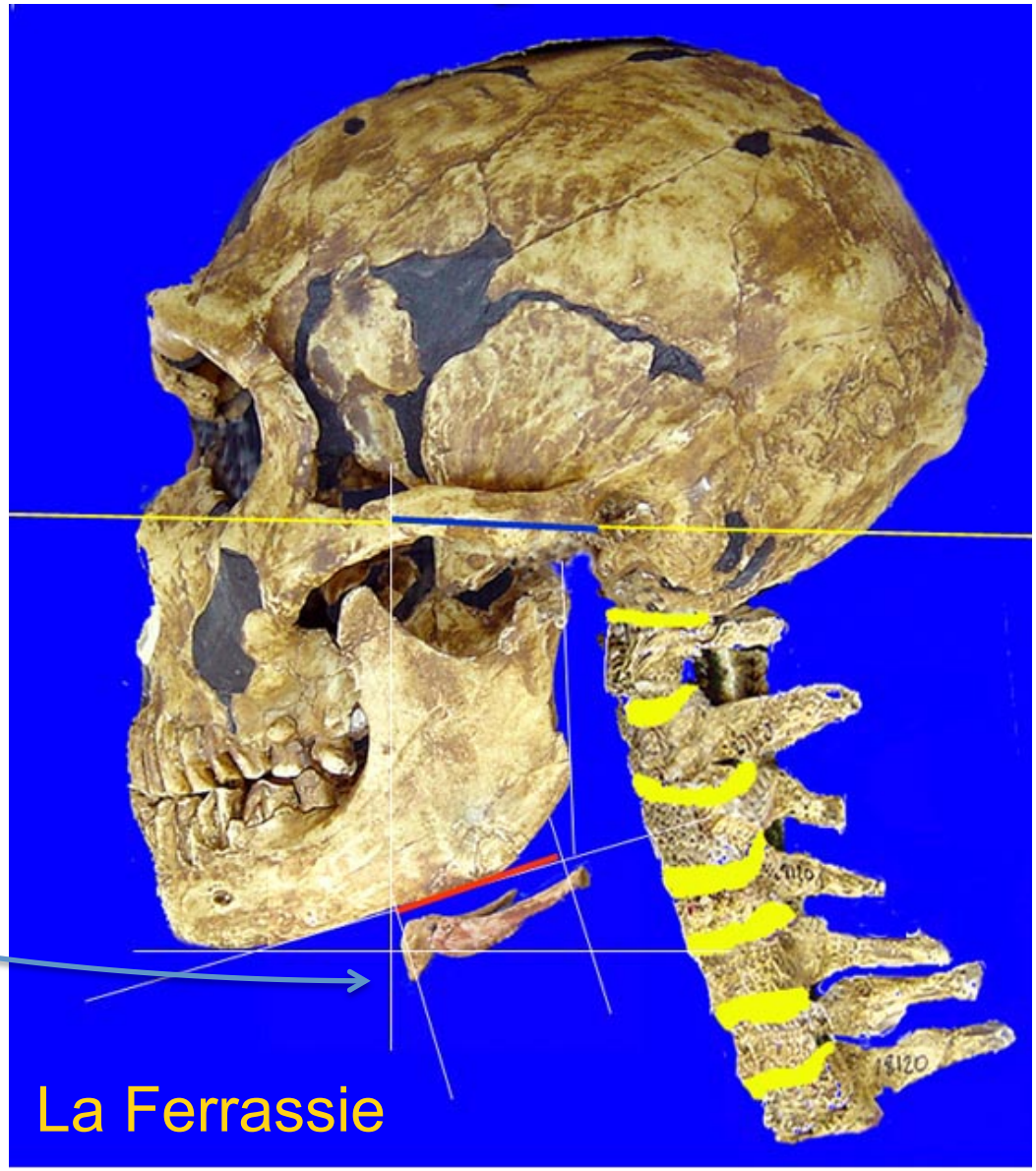
Fossils for speech?

- There is only one tiny bone in the vocal tract
  - The hyoid bone
  - And that is not connected to any other bone

- The rest is soft tissue
  - And almost doesn't fossilize at all



A reconstructed Neanderthal

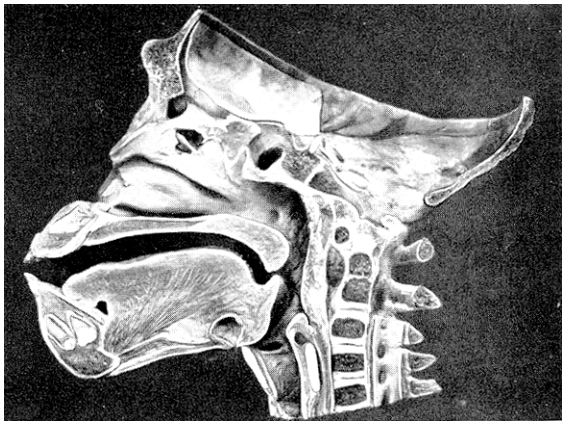


But this is Khebara!

Granat, J., Peyre, É. (2004) *Biométrie Humaine et Anthropologie* 22, 3-4, 141-163.
And Louis Jean Boë

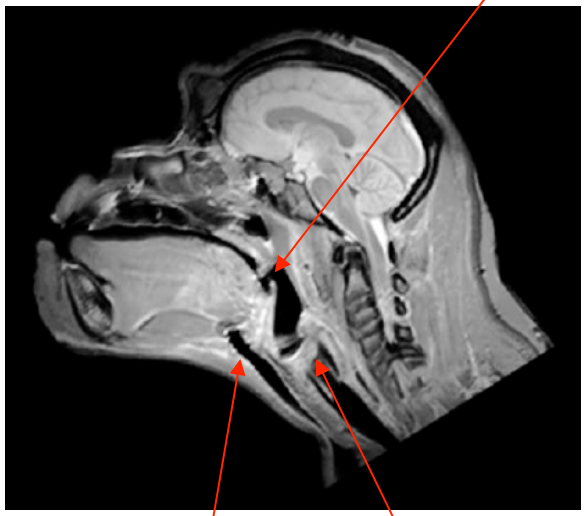
Vocal tract comparison

(Young) Gorilla



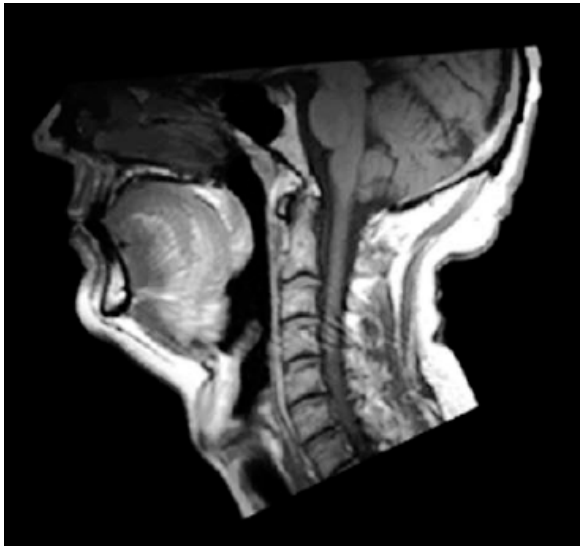
Negus 1949

Chimpanzee



Fitch 2000

Human Male



But what is the effect?

- By looking at the images, we cannot tell what the effect of these changes is
  - Or whether they are related to speech at all

- This is where computer models come in

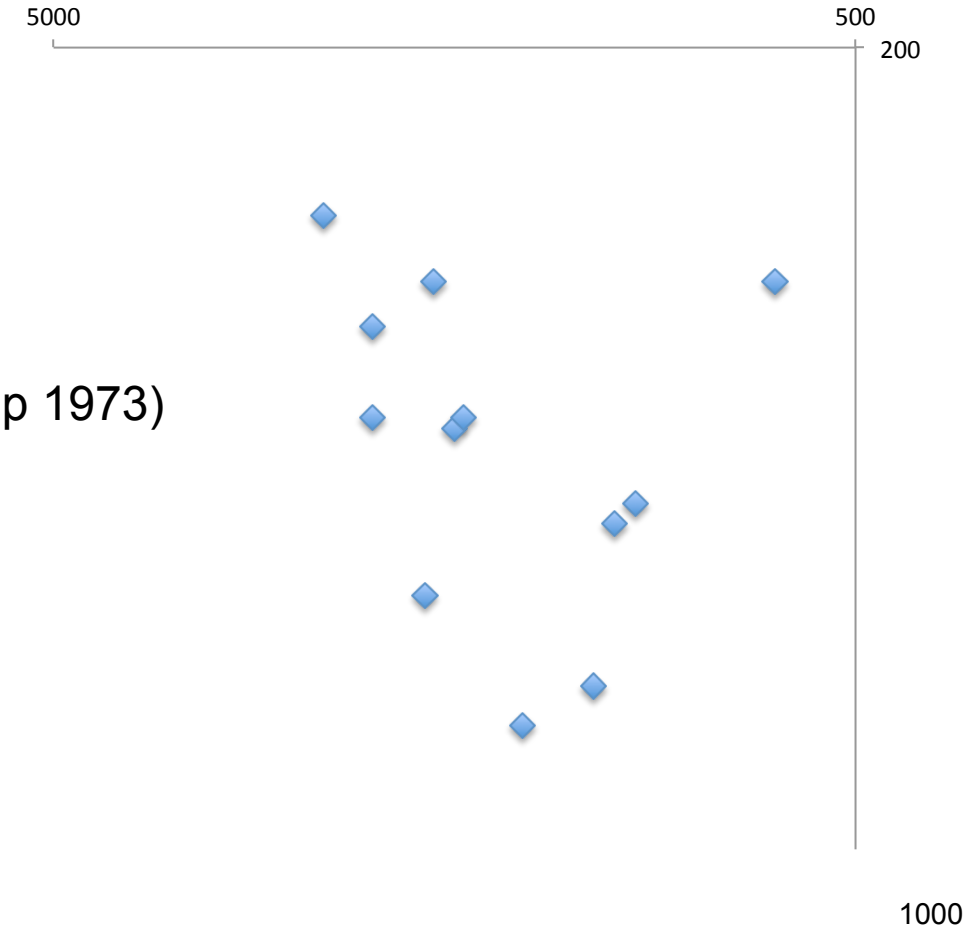
Four approaches

- Reconstructions of articulatory range
- Investigation of hypothetical anatomies
- Reconstruction of speech sounds
- Modeling population behavior

1: Reconstruction

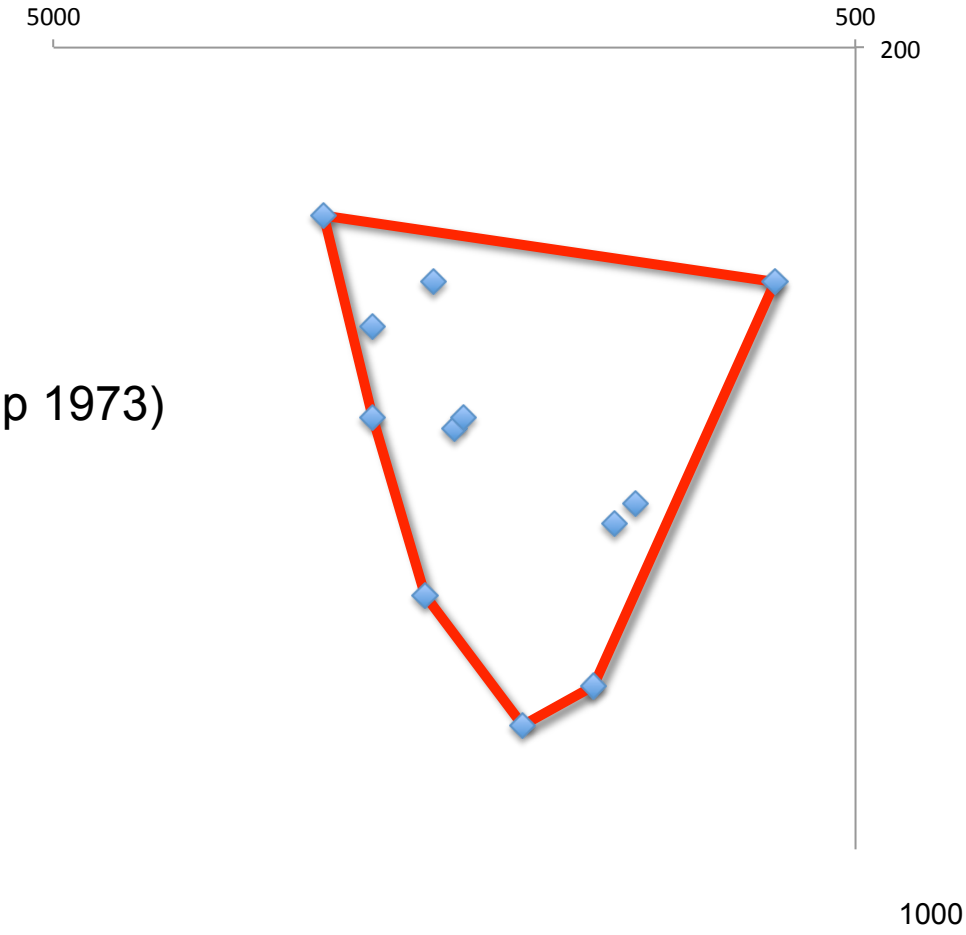
Maximal Vowel Space

Data from one
male speaker
of Dutch
(Pols & van Nierop 1973)

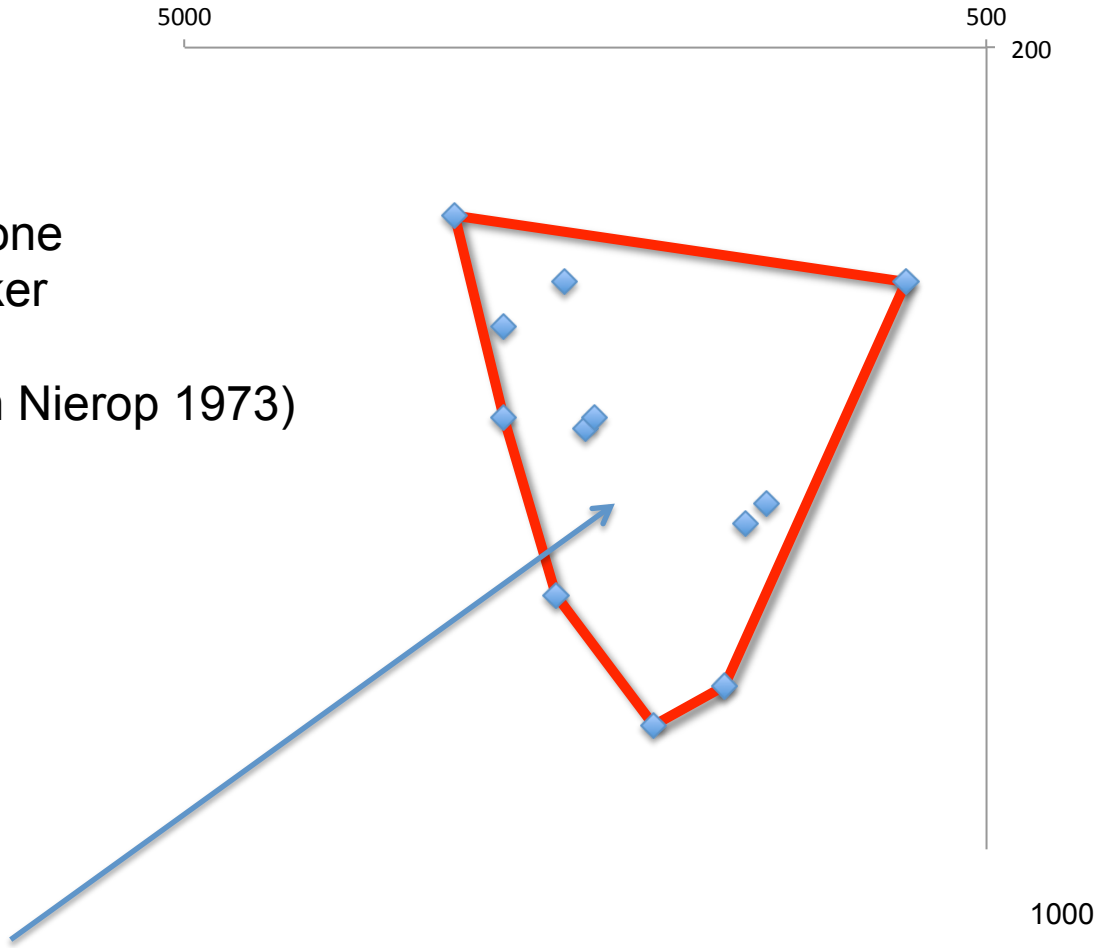


Maximal Vowel Space

Data from one
male speaker
of Dutch
(Pols & van Nierop 1973)



Maximal Vowel Space

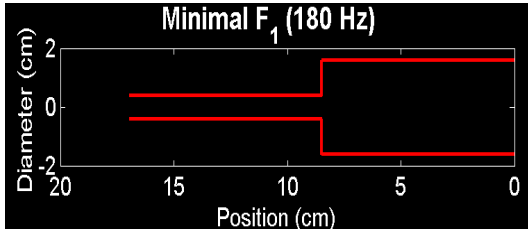


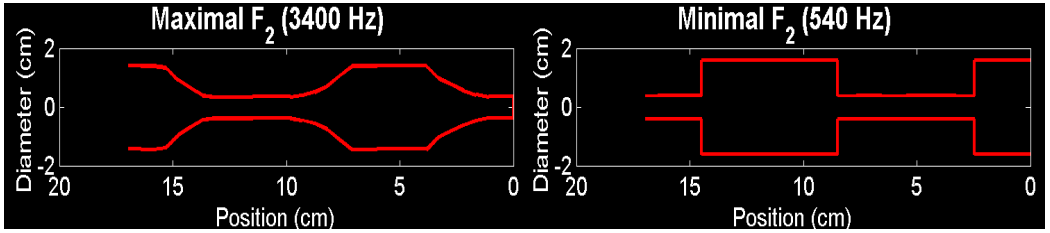
Data from one
male speaker
of Dutch
(Pols & van Nierop 1973)

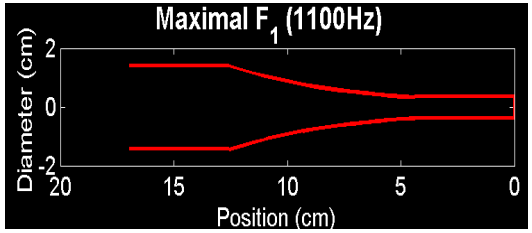
A speaker's maximal vowel space is the
size of the signal space that they can produce

How to make a maximal space

- Given
  - A tube of a certain length
  - A maximal area (anatomical limit)
  - A minimal area (turbulence)
- There are given maximally distinctive articulations
  - How can we achieve these?







Length = 16 cm
Minimal:maximal area = 1:8

Reconstructions of articulatory range

- The oldest kind of modeling (Lieberman et al. 1969)

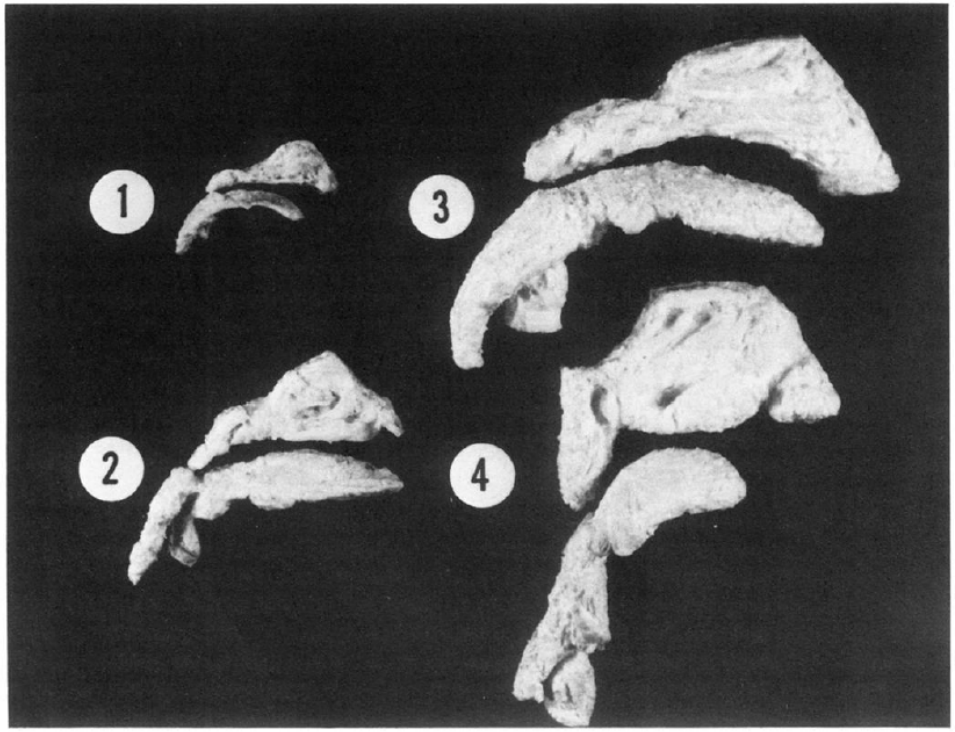


Figure 6. Casts of the nasal, oral, pharyngeal, and laryngeal cavities of (1) newborn human, (2) adult chimpanzee, (3) Neanderthal reconstruction, and (4) adult human.

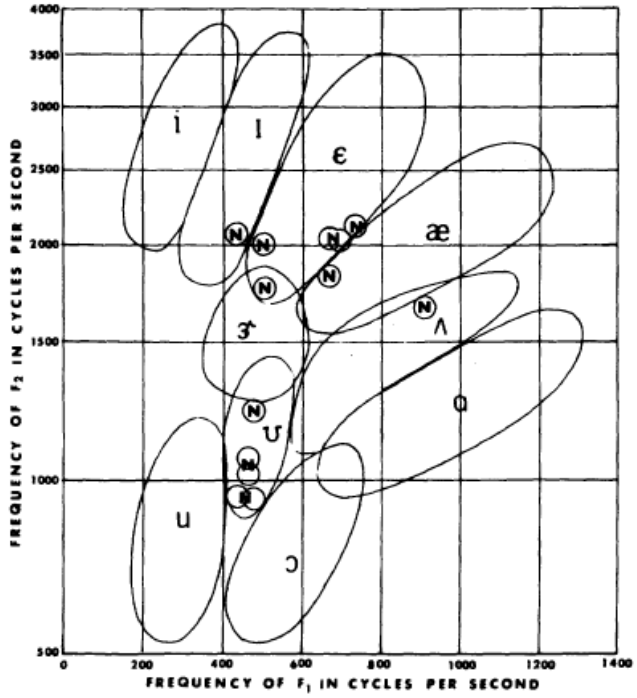
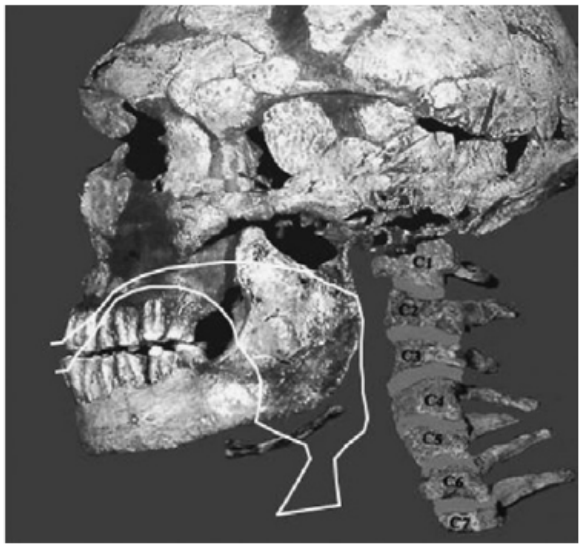


Figure 12. Plot of first and second formant frequencies for "extreme" vowels, data points (N), of reconstructed Neanderthal vocal tract (Lieberman and Crelin 1971).

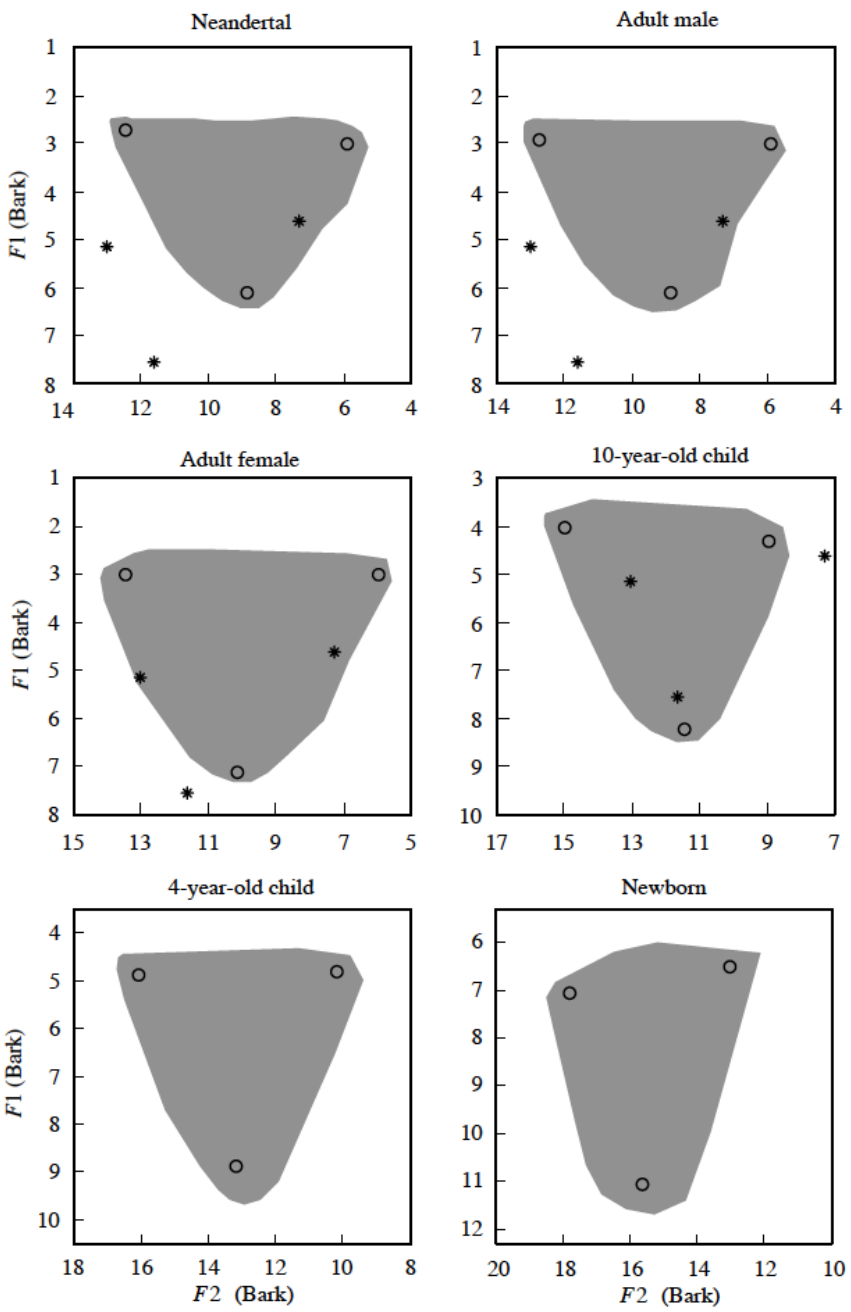
Lieberman et al. 1972

Boë et al.'s model

- A different model was made by Boë et al.
  - With a different (more modern) reconstruction
  - And with a very different computational model



Boë et al. 2007

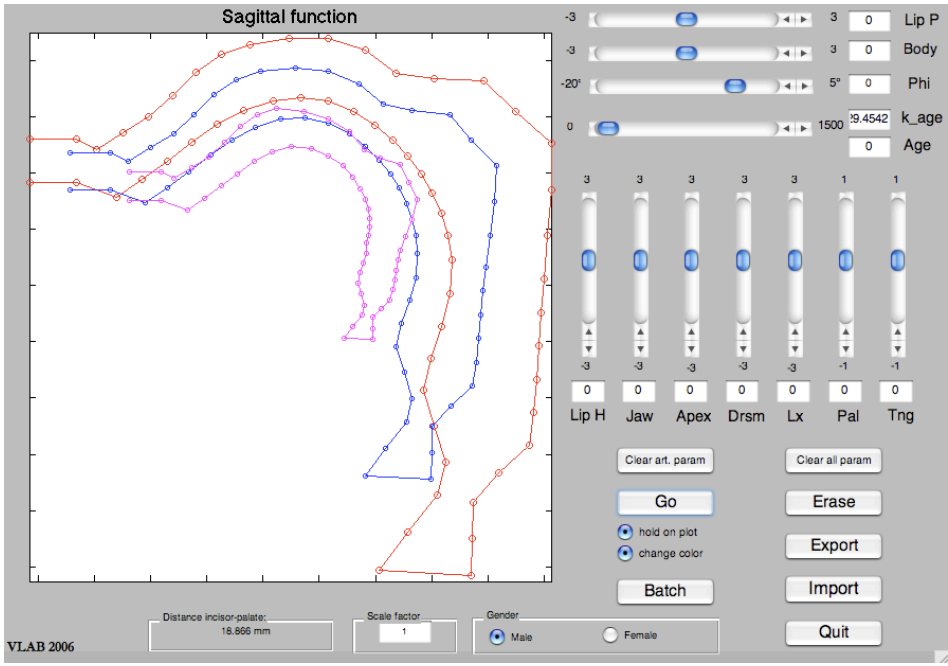


Boë et al. 2002

Figure 7. Maximal vowel spaces normalized by a perceptual scale in Bark and the values of F_1–F_2 for [i a u] proposed by Lieberman and Crelin for the Neandertal male (*), and for the Neandertal male, male adults and female adults estimated in this study (○).

VLAB: Boë's model

- Vocal tract laboratory
  - Formerly VLAM – Variable Linear Articulatory Model
- The dimensions of different parts of the tract can be controlled
  - Based on X-ray and developmental data



From a presentation by Boë

Controversy (1)

- In order to make such a model, you have to make many assumptions
  - In which the anatomical reconstruction is the "easiest" part
  - But you need to make guesses about physiology and control as well

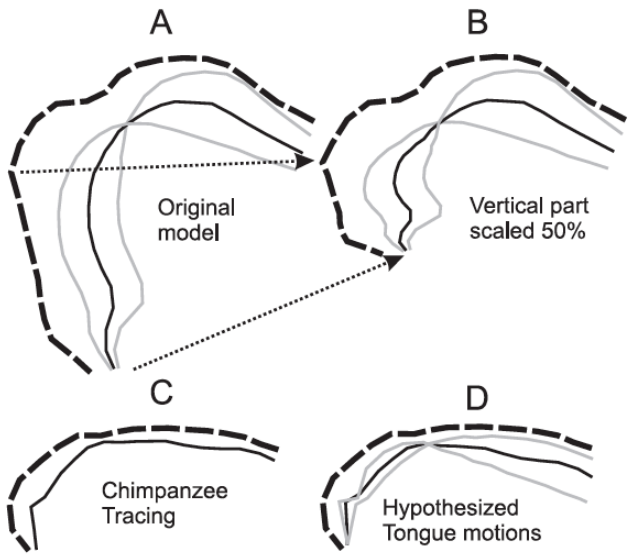


Figure 2 Principle underlying Boë's (1999) adaptation of Maeda's (1990) model. The vertical part and the horizontal part of Maeda's vocal tract can be scaled independently. Outline A shows Maeda's original model and the effect of manipulating his first principal component (see Figure 1). The outline B shows what happens when the vertical part is scaled 50%. The tongue root is deformed unrealistically, because the motions of the original tongue are preserved despite drastic changes in actual muscle angles. Outline C shows a tracing of a chimpanzee vocal tract (Fitch, 2000a, Figure 1), and outline D the hypothesized actions (corresponding to Maeda's first principal component) it could perform with its tongue in the resting (high tongue-root and larynx) position.

de Boer & Fitch 2010, Adapt. Beh. 18: 36-

Controversy (2)

- In addition, the debate has been muddled by confusing two independent questions
  - Did Neandethals speak?
    cannot really be answered
  - Did the vocal tract evolve for speech?
    I would argue this can be answered

2: Hypothetical anatomies

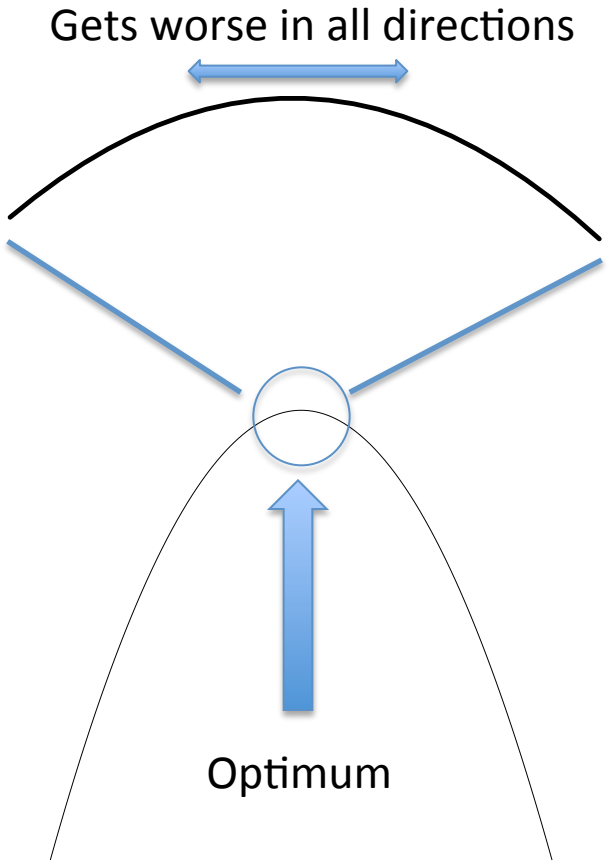
Hypothetical anatomies

- There are too many unknowns in vocalization
  - Any reconstruction can be disputed

- In order to progress we must investigate hypothetical anatomies
  - But why investigate something that doesn't exist?

Optimality Theory

- We can ask ourselves:
  - Did the modern vocal tract evolve for speech?
  - If it is optimal for vocalization, then most likely: yes (Parker and Maynard Smith 1990)
  - To establish optimality, explore neighboring anatomical configurations
  - Useful, even if these configurations are hypothetical



The Mermelstein model

- Based on geometrical/ anatomical modelling of an X-ray film

- Developed at Bell laboratories

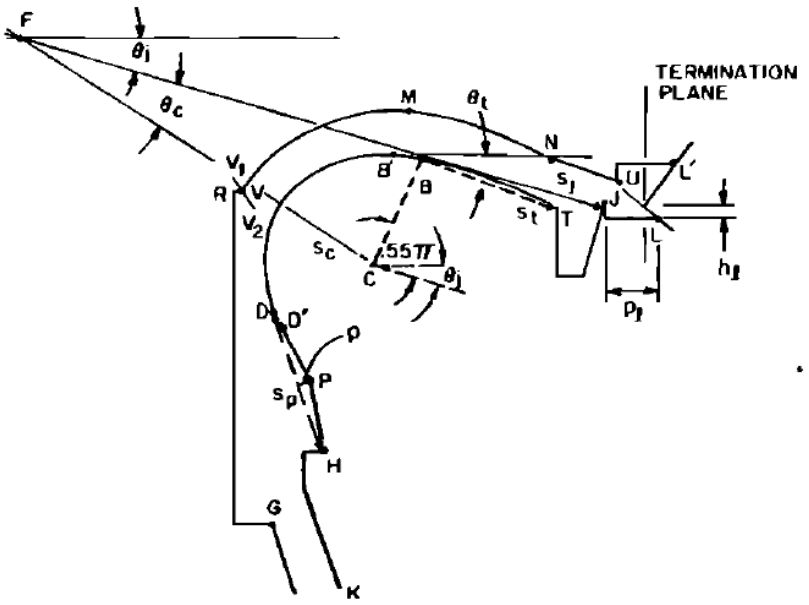
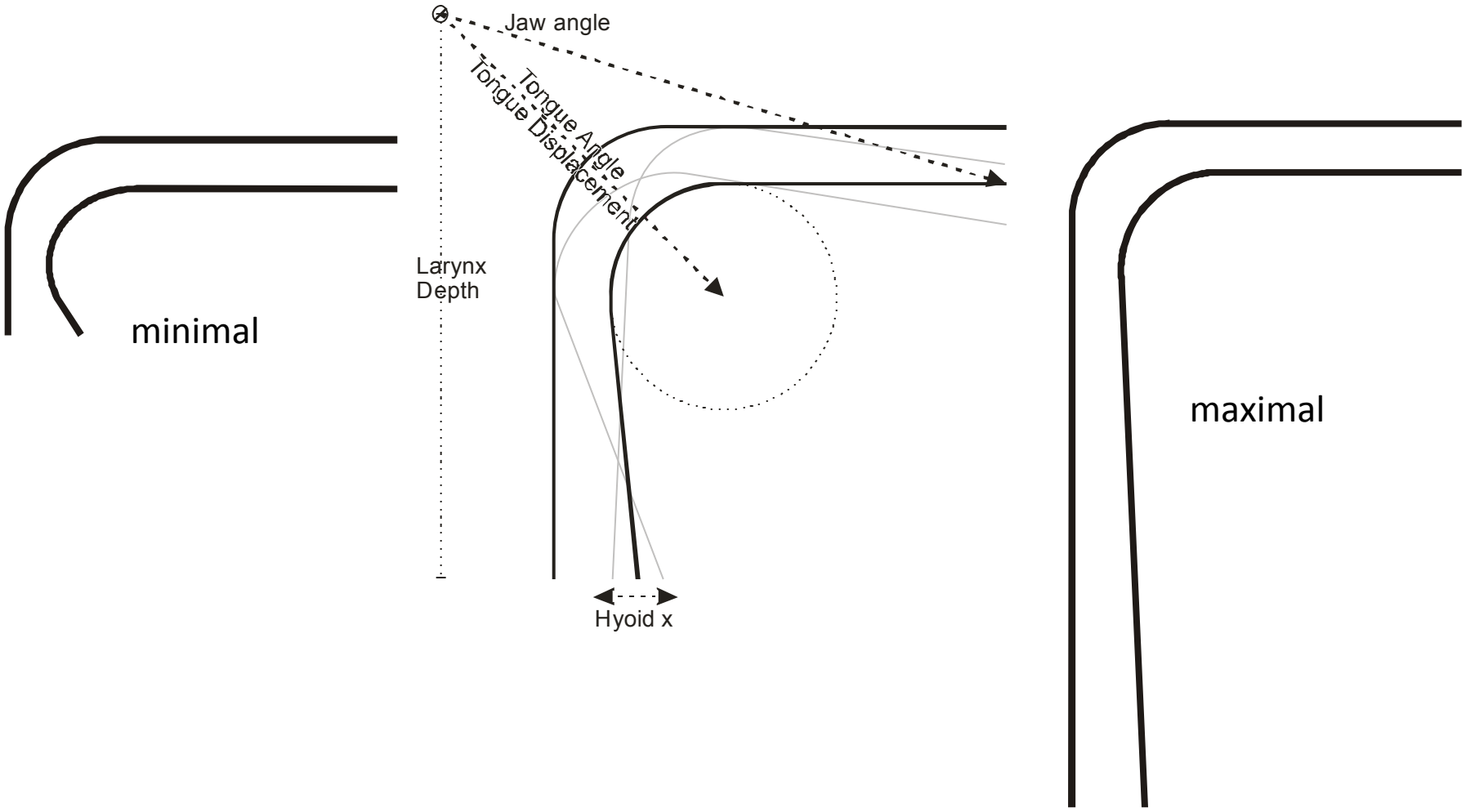


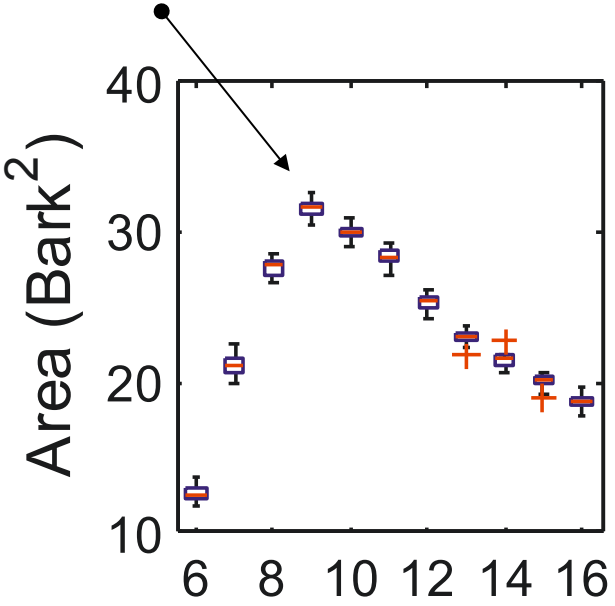
FIG. 1. Model-generated vocal-tract outline.

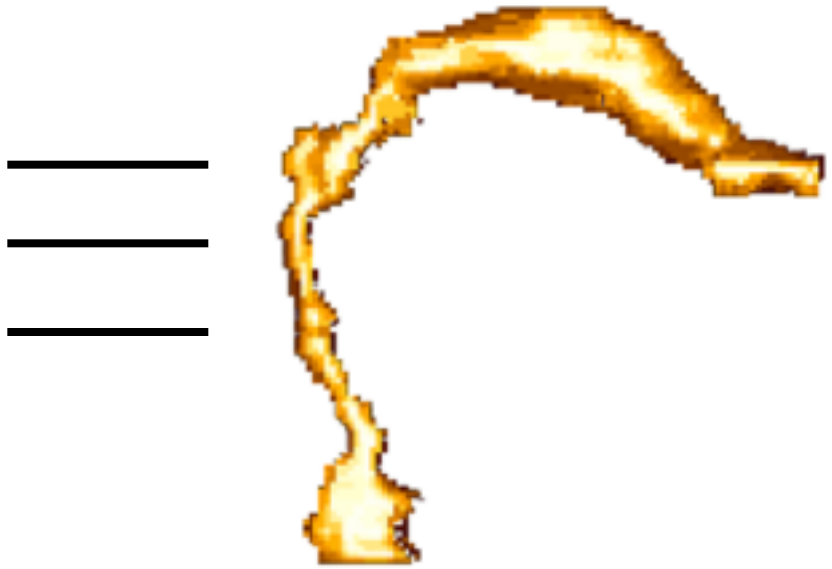
Virtual vocal tracts



Optimality

Optimal at 9 cm larynx depth





Female vocal tract

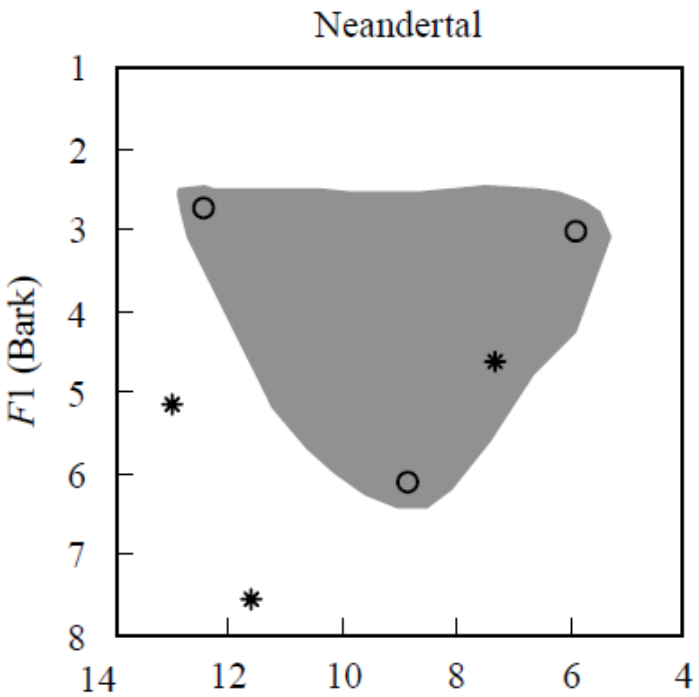
de Boer (2010) *J. Evol. Psych*. **8**: 351-

Story et al. 1998

3: Reconstructed sounds

Reconstructed sounds (1)

- In reconstructed and hypothetical anatomies
  - Range of sounds is not necessarily a subset of present-day speech sounds
  - OK, if anatomy is quali-tatively comparable
  - But even then, the range may be very different



Boe et al. 2002

Air sacs

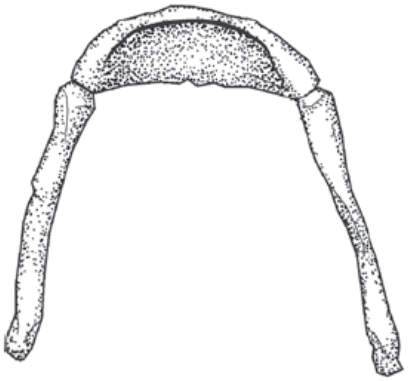
Orang Utan

Brown Howler monkey

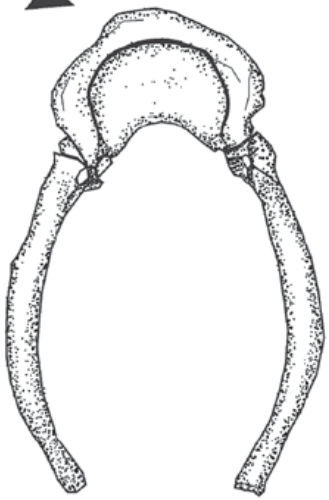
Siamang

Human

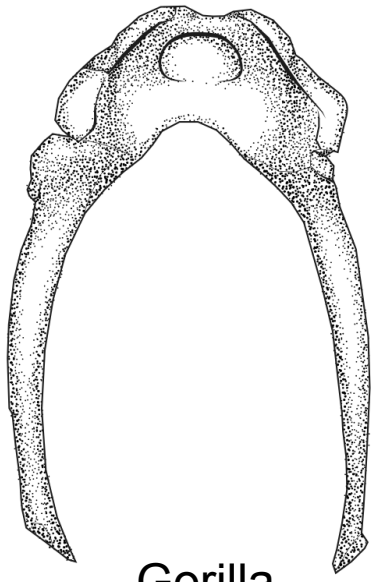
Hyoid bones



Human

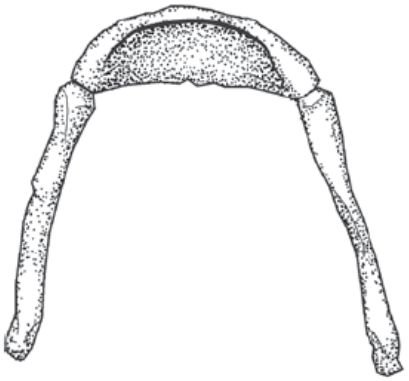


Chimpanzee

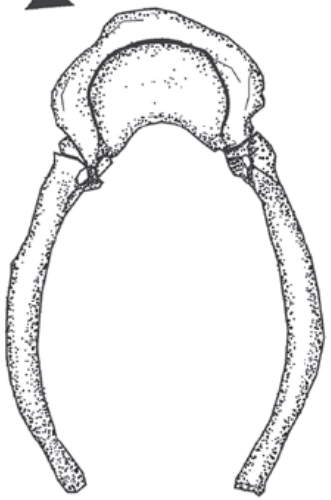


Gorilla

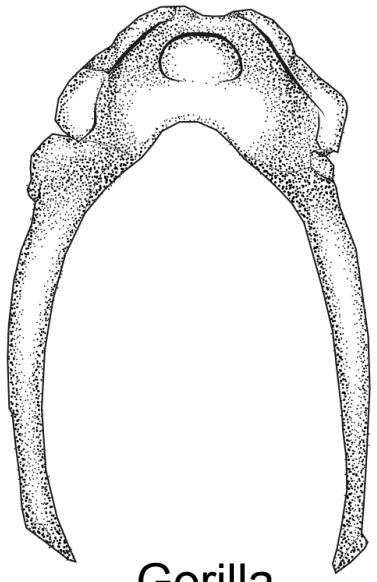
Hyoid bones



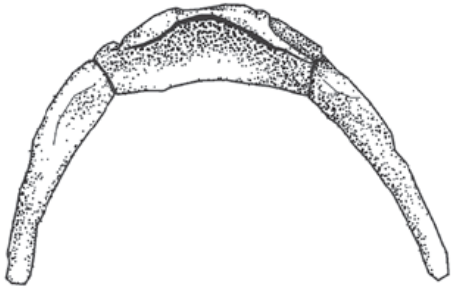
Human



Chimpanzee



Gorilla



Neanderthal

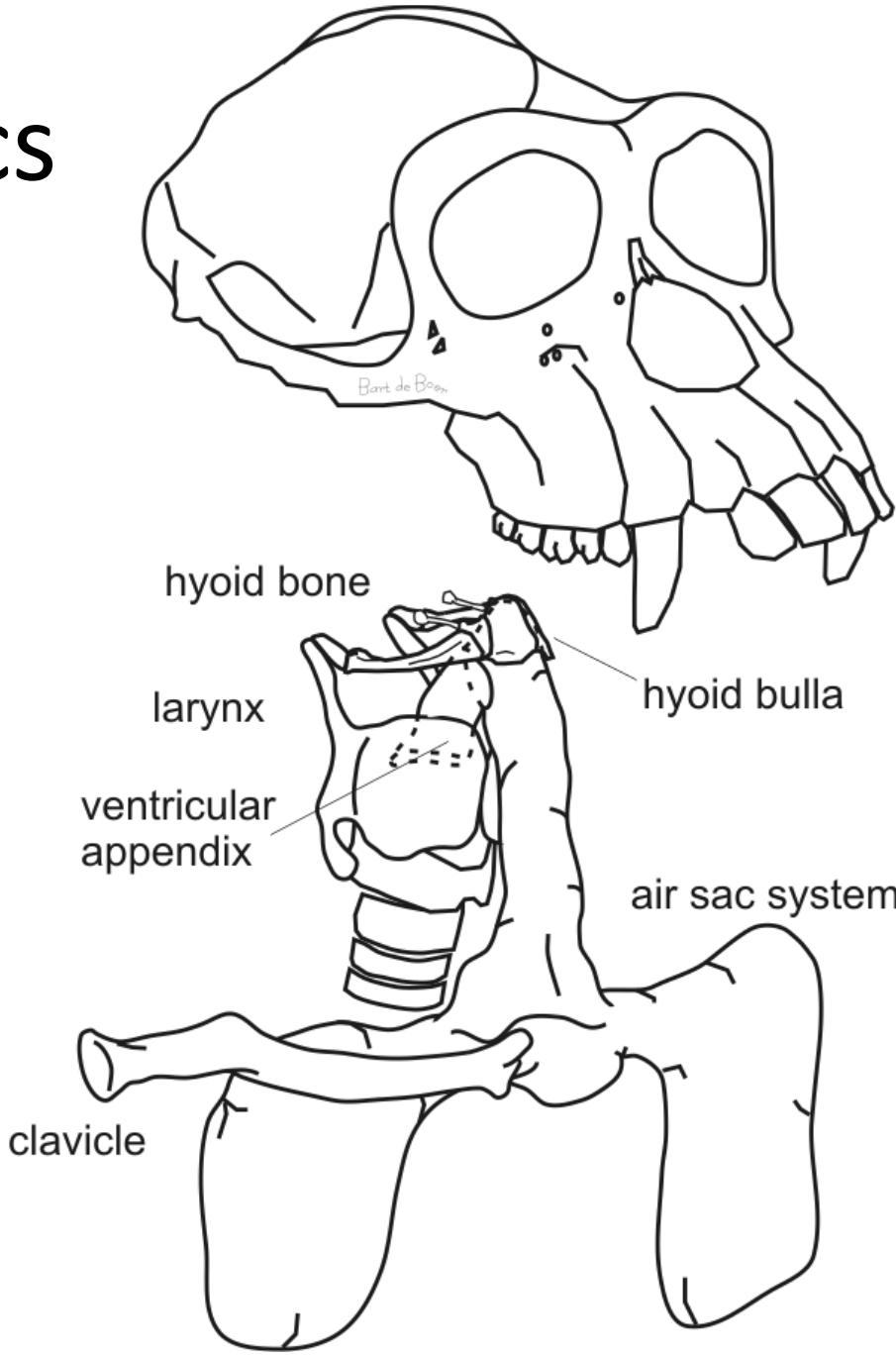
Dikika baby, *A. afarensis*
(Alamseged et al. 2006)

Australopithecus afarensis

Hyoids and air sacs

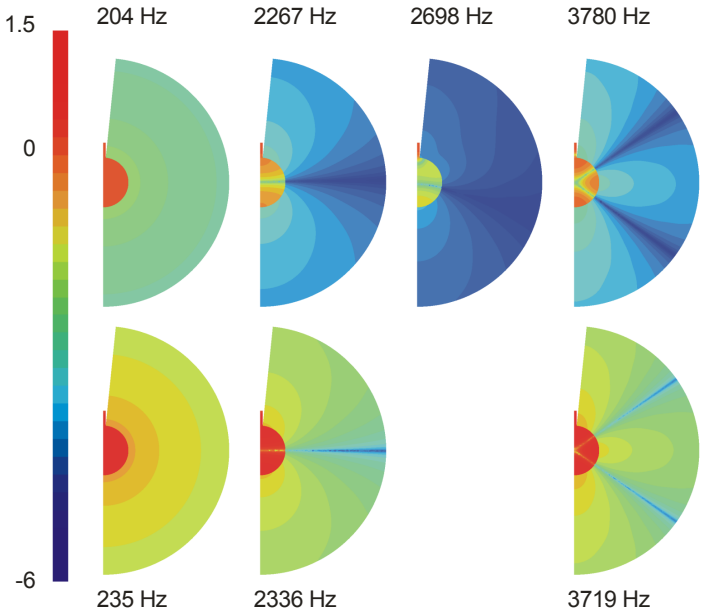
- The hyoid bulla helps to keep the connection to the air sac open (?)

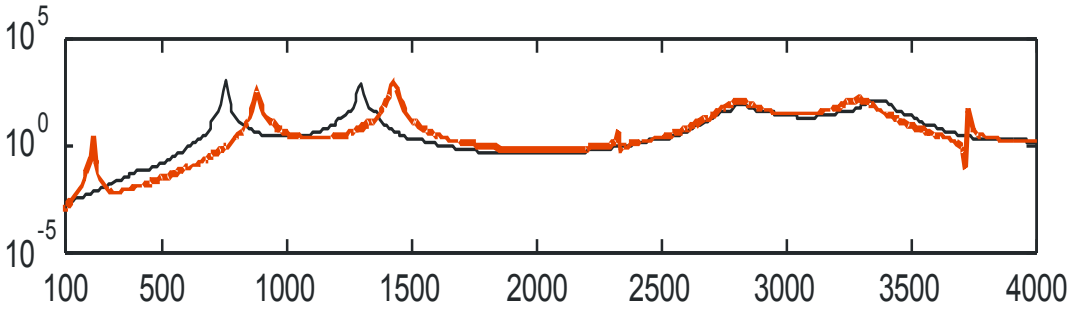
- Australopithecines had air sacs, Neanderthals didn't
  - But what does this have to do with speech?



Reconstructed sounds (2)

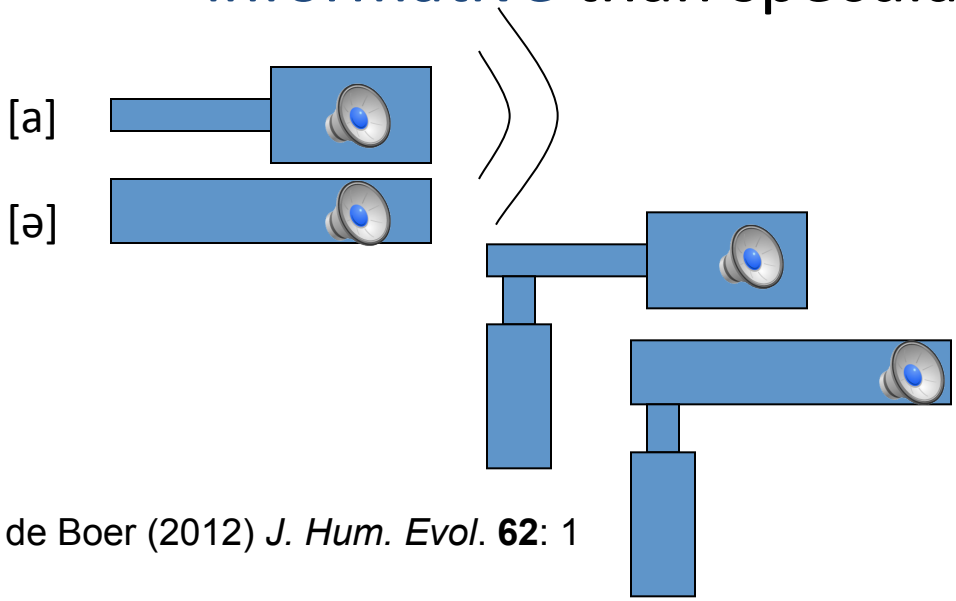
- How would sounds with non-human like formant patterns have been perceived?
  - For example, an air sac causes qualitatively different formant patterns

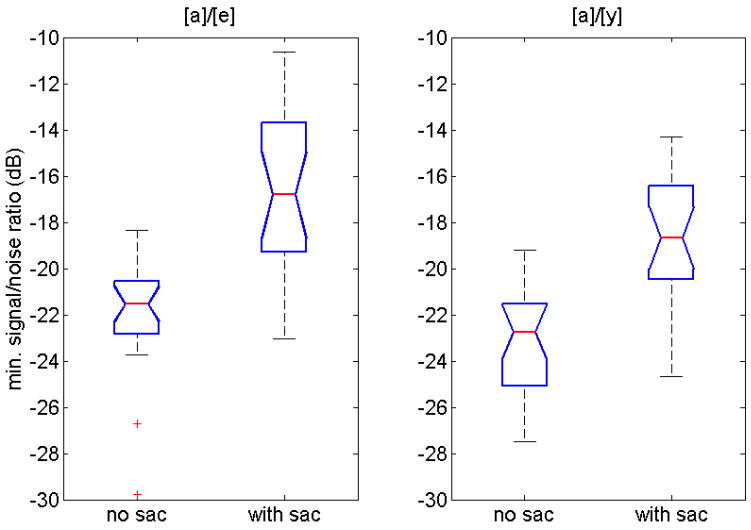




Reconstructed sounds (3)

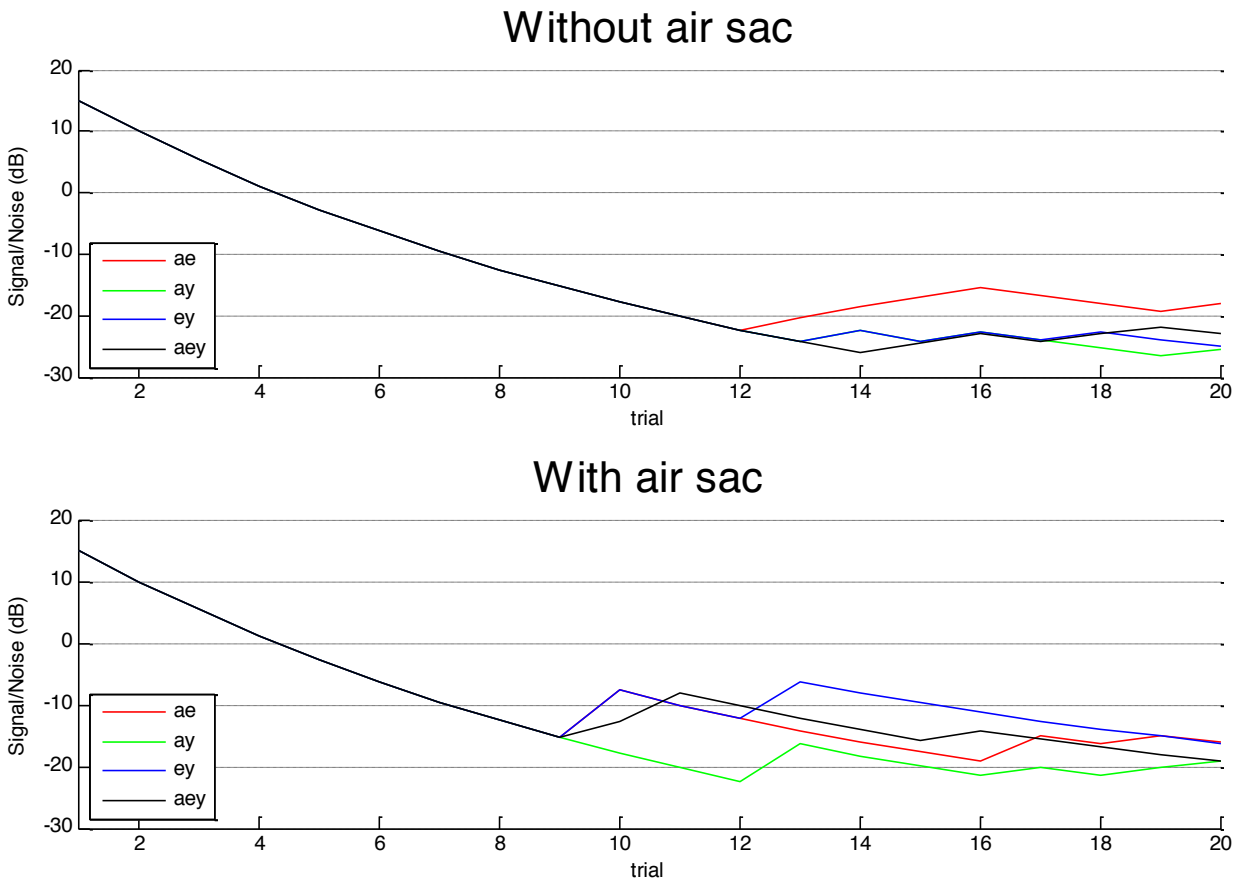
- Hearing appears to be more conservative
  (although there is discussion about how conservative exactly)
- But even if there are small differences
  - Experiments with modern subjects are more informative than speculation





de Boer (2012) *J. Hum. Evol.* **62**: 1

Example trial



3: Models of Cultural Change

Computational models

- Self-organization in vowel systems

- Emergence of phonemic structure

The aim of the study

- Explain universals of vowel systems
  - Why are do certain (combinations of) vowels occur more often than others
    (acoustic distinctiveness)
  - *How* does the optimisation take place?
- Hypothesis
  - Self-organisation in a population under constraints of production, perception, learning causes optimal systems to emerge
- Model
  - Agent-based model
  - *Imitation games*


JPHO 20000125 PROOFTYPE: TYP

Journal of Phonetics (2000) 28, 000–000
doi:10.1006/jpho.2000.0125
Available online at http://www.idealibrary.com on IDEAL®

Self-organization in vowel systems

Bart de Boer*

AI-lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium

Received 25th September 1998, and accepted 9th October 2000

This paper presents a computer simulation of the emergence of vowel systems in a population of agents. The agents (small computer programs that can play imitation games with each other) are based on a realistic model of synthesis, a model of human perception and the ability to imitate and learn sounds they hear. It is shown that due to the interactions between the agents and due to self-organization, realistic vowel repertoires emerge. This happens under a large number of different parameter settings and therefore seems to be a very robust phenomenon. The emerged vowel systems show remarkable similarities with the vowel systems found in human languages. It is argued that self-organization should be taken seriously as an explanation for many universal properties of human languages and that innate predispositions are probably not necessary to explain the universal tendencies of human vowel systems.

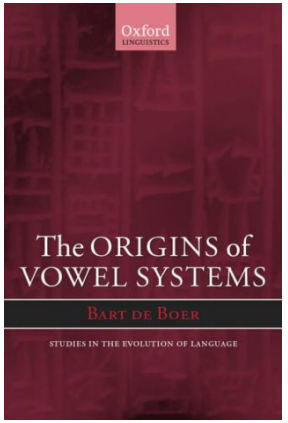
1. Introduction

The sound systems of human languages show remarkable regularities. Although humans are able to produce and perceive many different speech sounds, languages do not use random subsets of these sounds. In UPSID, the UCLA Phonological Segment Inventory Database (Maddieson, 1984; Maddieson & Precoda, 1990) which now contains 451 languages, 921 different speech sounds are found. However, the average number of phonemes that is used in the languages in this sample lies between 20 and 37 (Maddieson, 1984). The minimal number of phonemes in any language in the sample is 11 in Rotokas (Fischer & Fischer, 1957) and Pirahã (Everett, 1982) and the maximum number is 141 in !Xũ (Snyman, 1975). This means that human languages generally use only a small subset of the available speech sounds.

Within these subsets, some speech sounds occur more often than others. Some sounds, such as [a] in 87% of the languages in UPSID_{317} ([m] 94%; or [k] 89%) occur almost universally while others, such as [ɛ] (2%) [ɑ] (1%; or [ʃ] (1%) occur only rarely. Moreover, phoneme inventories tend to exhibit certain symmetries. If, for example, a repertoire contains an [o] (in 36% of the languages of UPSID_{317}) it is almost 5 times more likely to contain an [e] (in 41% of the total sample) than if it does not contain [o].

* E-mail: bart@arti.vub.ac.be

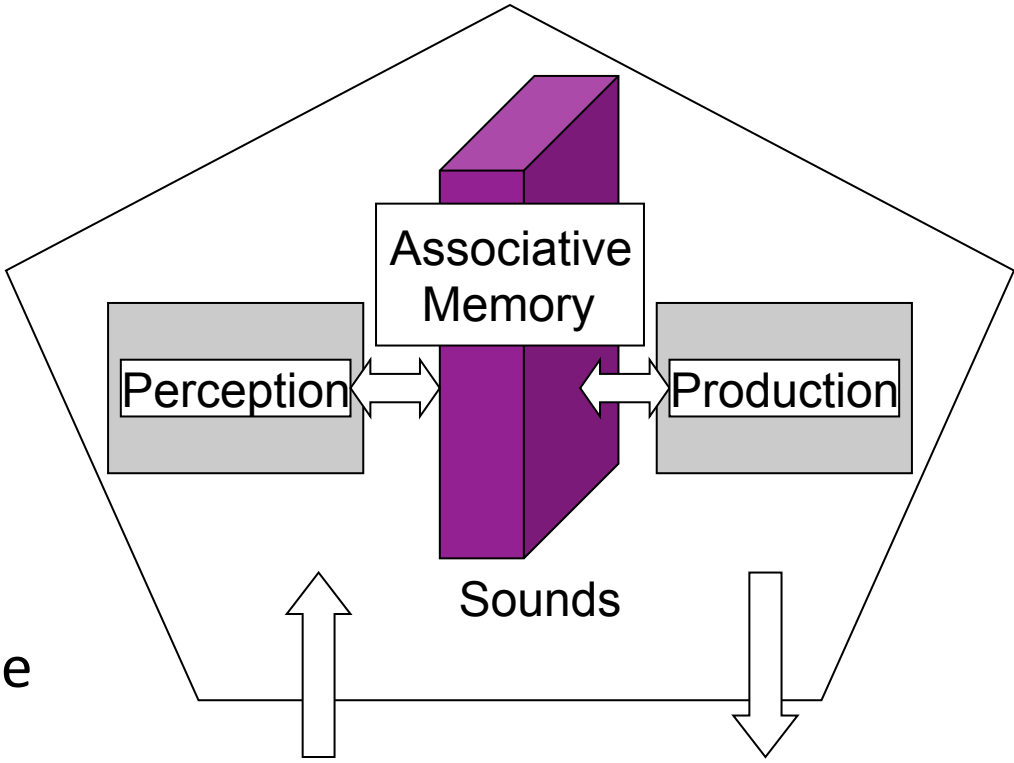
0095-4470/00/000000 + 00 $35.00/0



Architecture

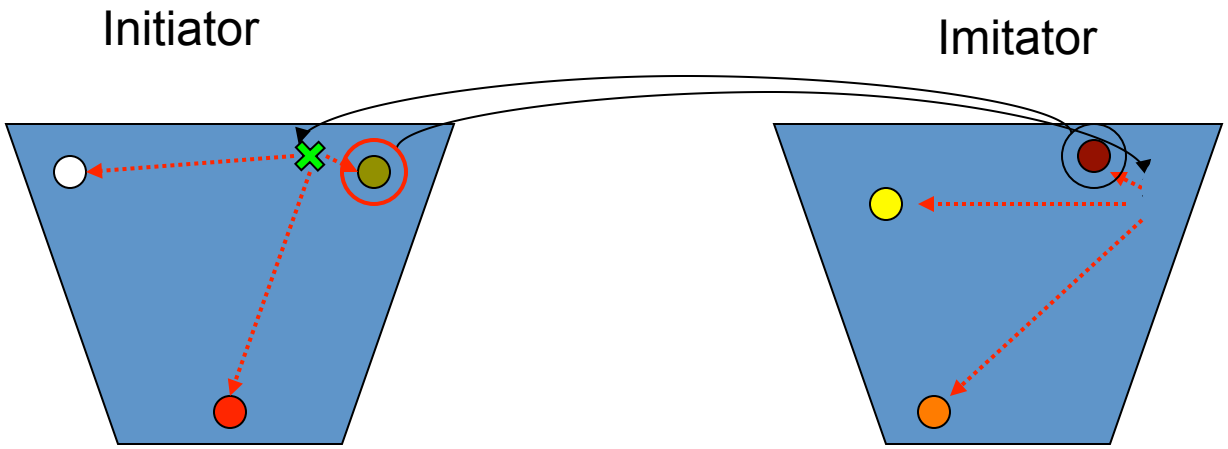
For vowels:

- Realistic production
  articulatory synthesiser
  (Maeda, Valleé)
- Realistic perception
  Formant weighting
  (Mantakas, Schwarz, Boë)
- Learning model
  Prototype based associative memory

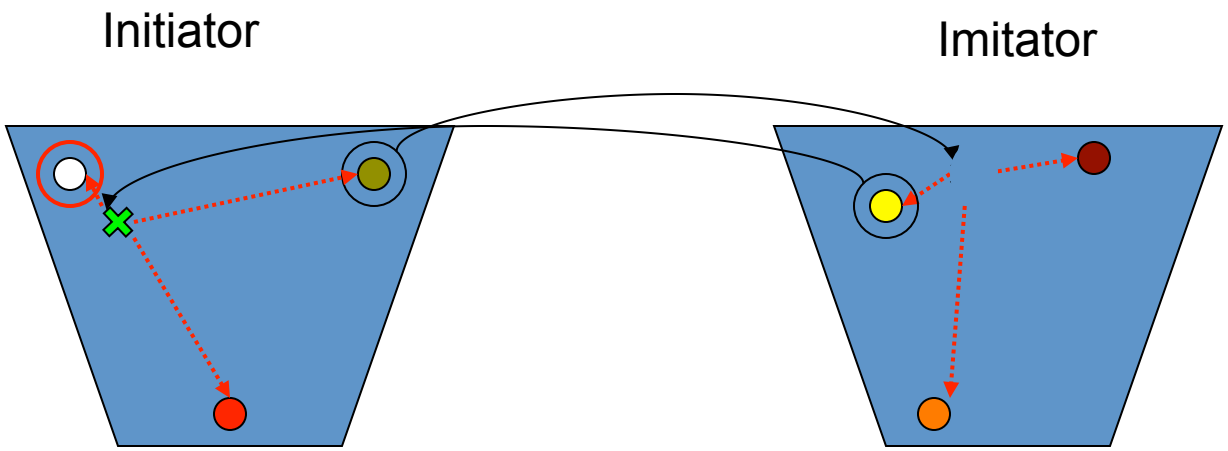


The interactions

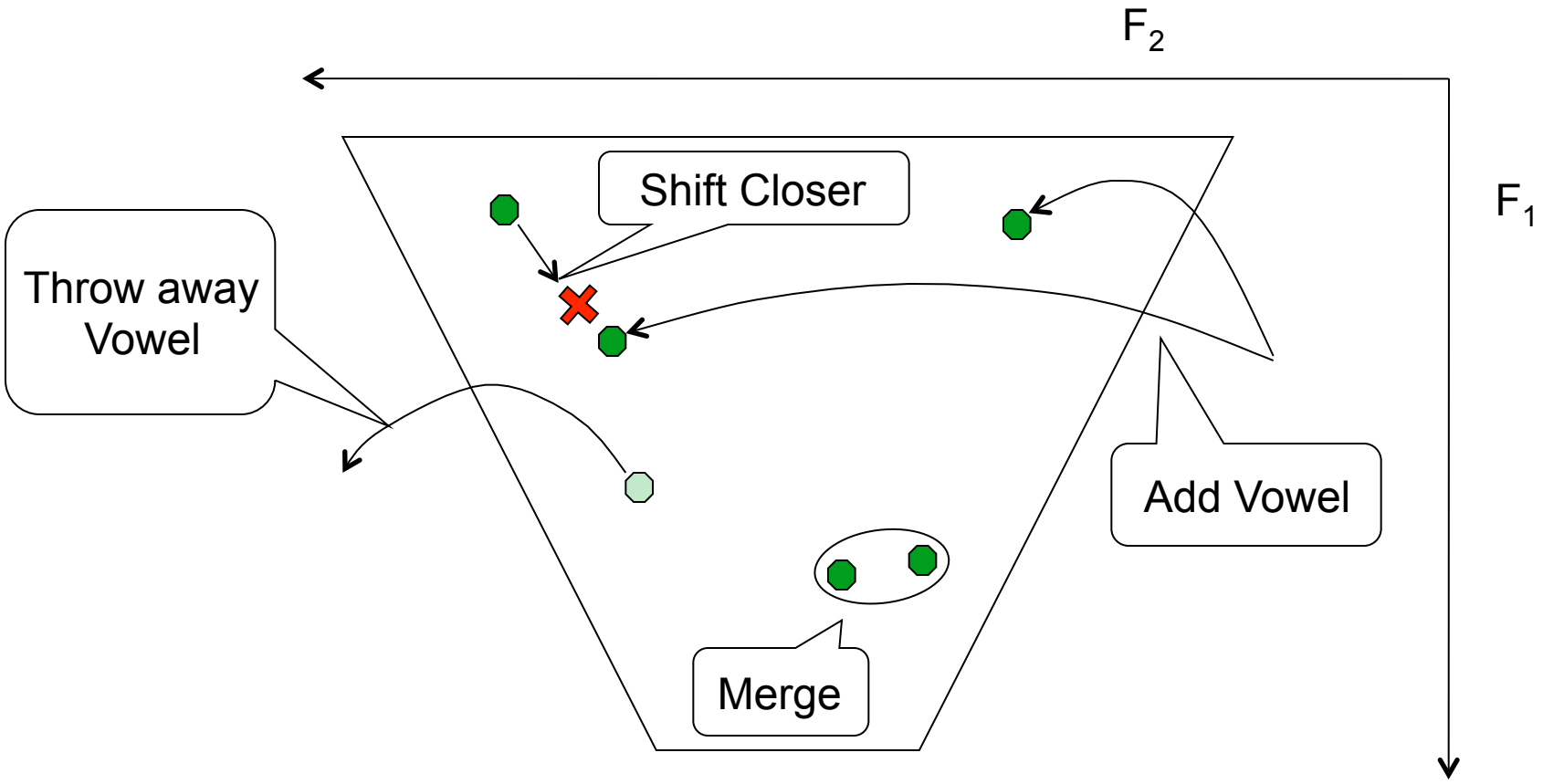
- Imitation with categorical perception
  - Humans hear speech signals as the nearest phoneme in their language (?)
- Correctness of imitation depends not only on the signals used, but also on the agents' repertoires



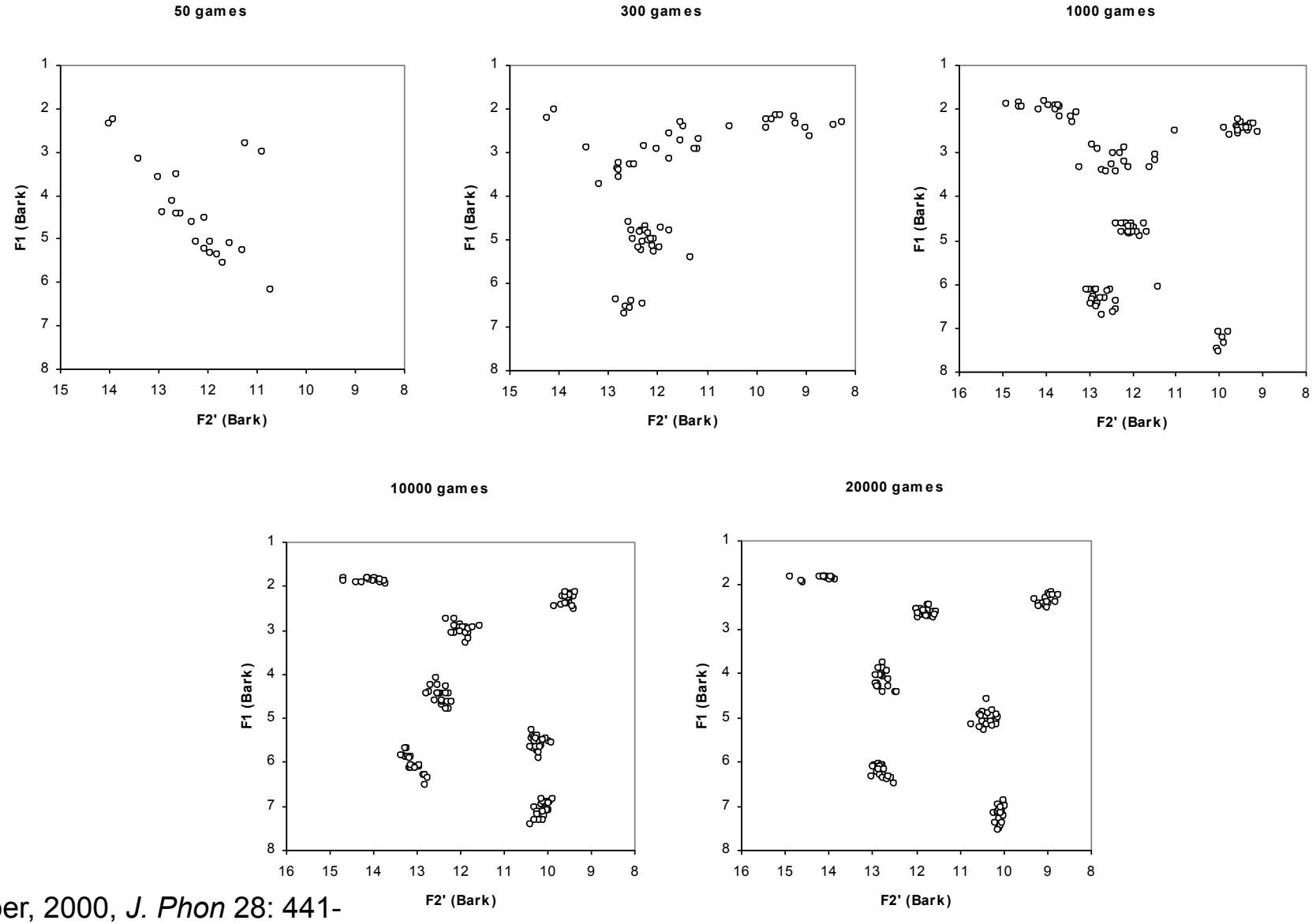
Imitation failure



Reactions to imitation game



The emergence of vowel systems



de Boer, 2000, *J. Phon* 28: 441-

Combinatorial speech

- Human speech is combinatorial
- Ape speech is not
- For productive use, special abilities appear to be necessary
- But how would that evolve if the language itself is not already combinatorial?
- Chicken and egg problem of language evolution

Hypothesis

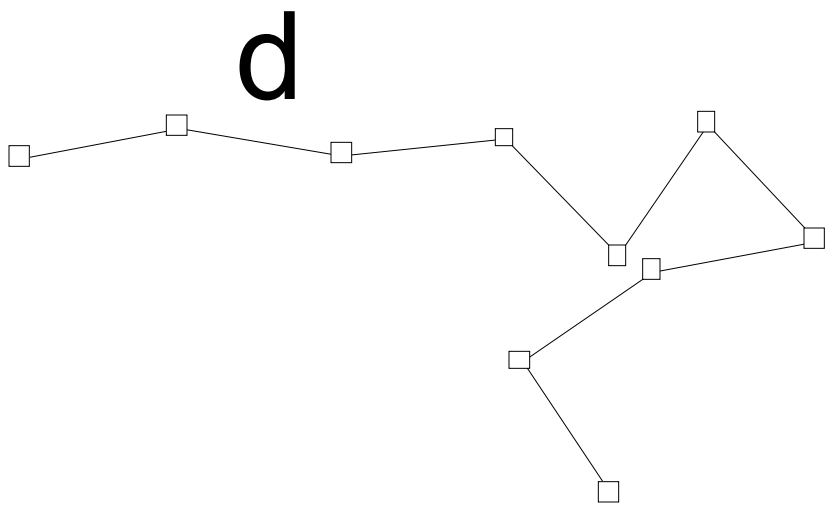
- Optimizing distinctiveness of trajectories that are stretched out in time results in superficial combinatorial structure
- Can occur when there is a growing repertoire of utterances
- Productive use follows later
- Work together with Jelle Zuidema

The basic algorithm

- Repeat until convergence:
  1. Select a random trajectory from the repertoire
  2. Make a random change to this repertoire
  3. Check whether total distance increases
  4. If yes, keep change

Trajectories

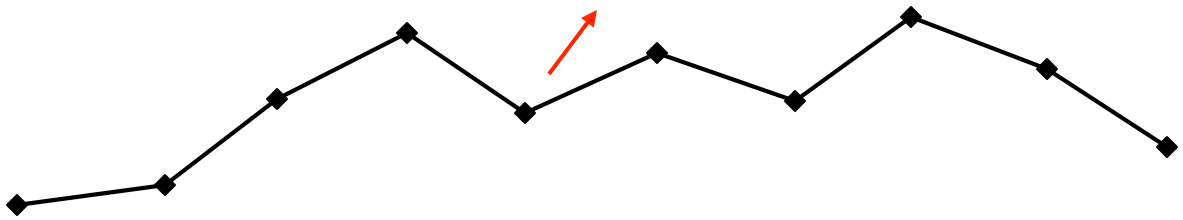
- Points on a trajectory are equidistant, but can have any angle with respect to each other
- Trajectories exist in a square 2-D space



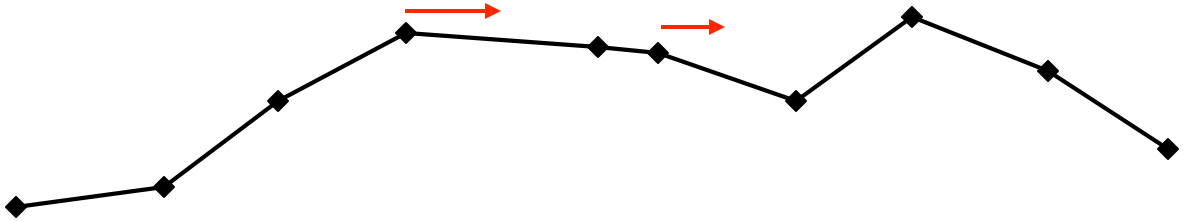
Shifting a trajectory

- Randomly select a point and shift it over a random vector
- Iteratively shift preceding and following points on trajectory in the direction of neighboring points

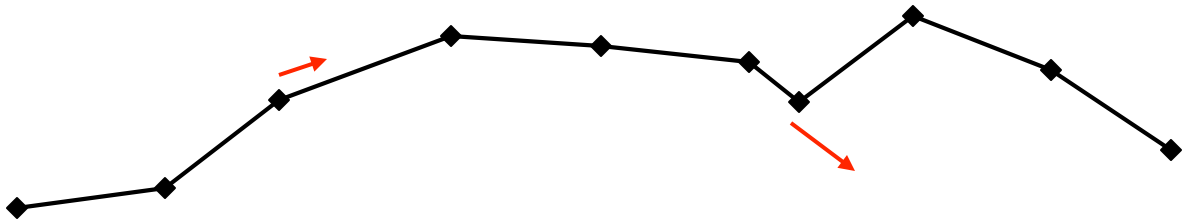
Shifting-example (1)



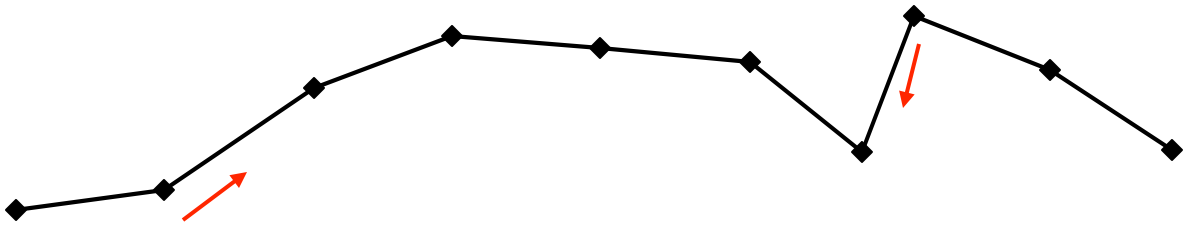
Shifting-example (2)



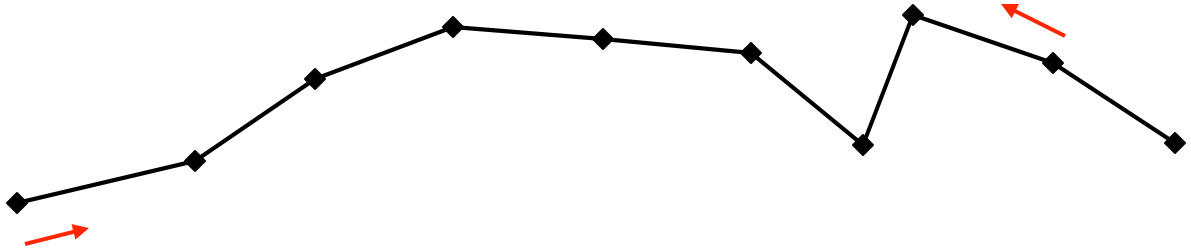
Shifting-example (3)



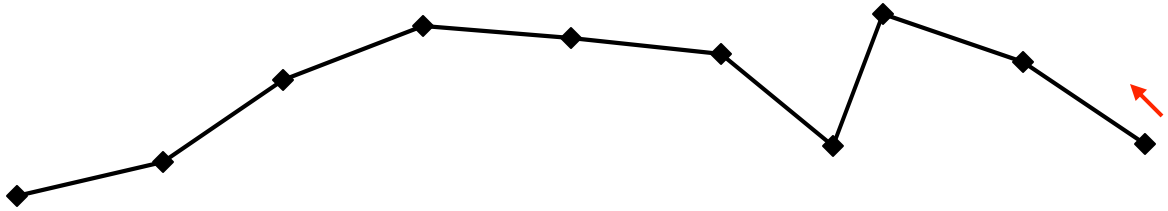
Shifting-example (4)



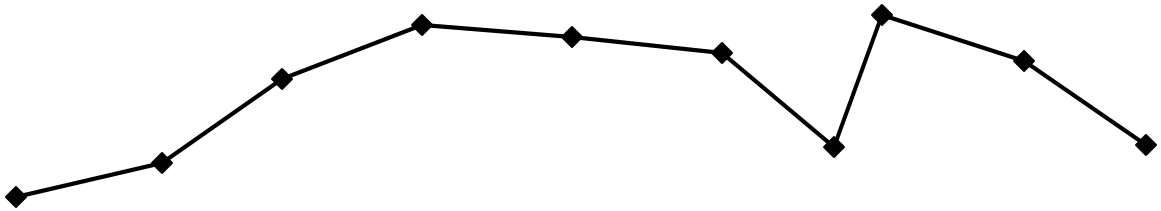
Shifting-example (5)



Shifting-example (6)

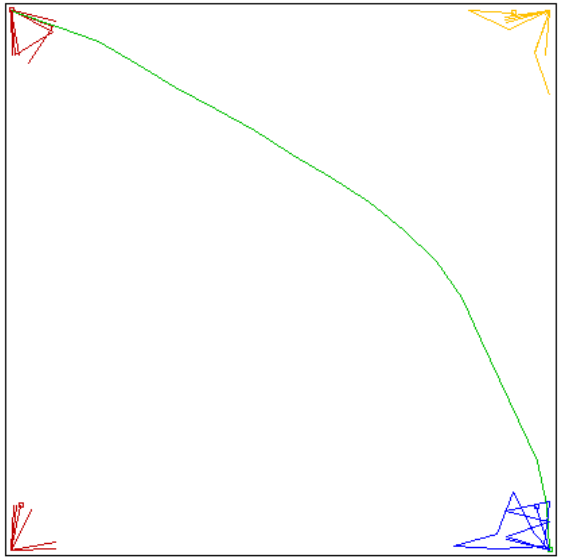


Shifting-example (7)

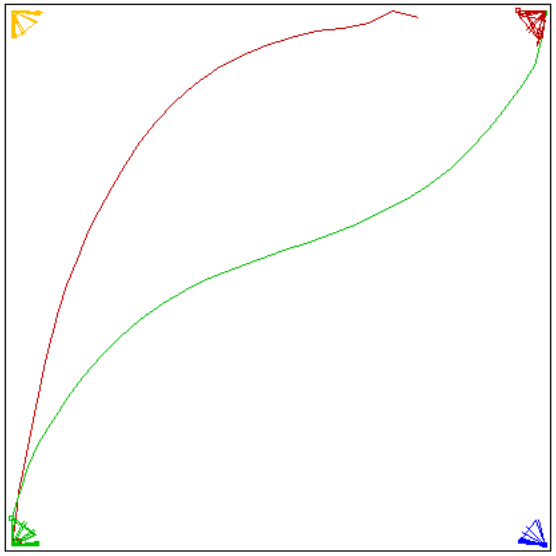


Results

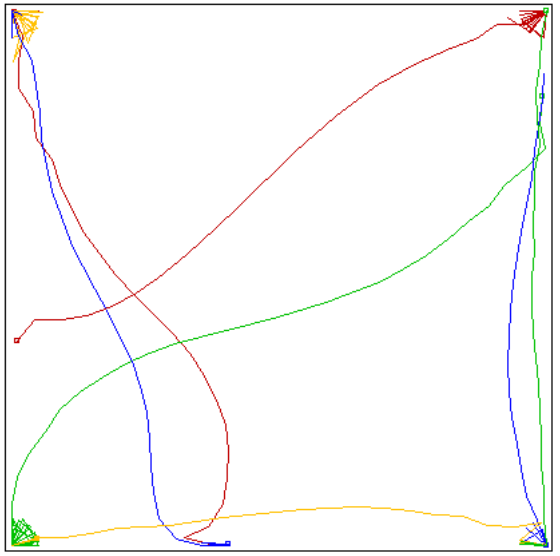
5



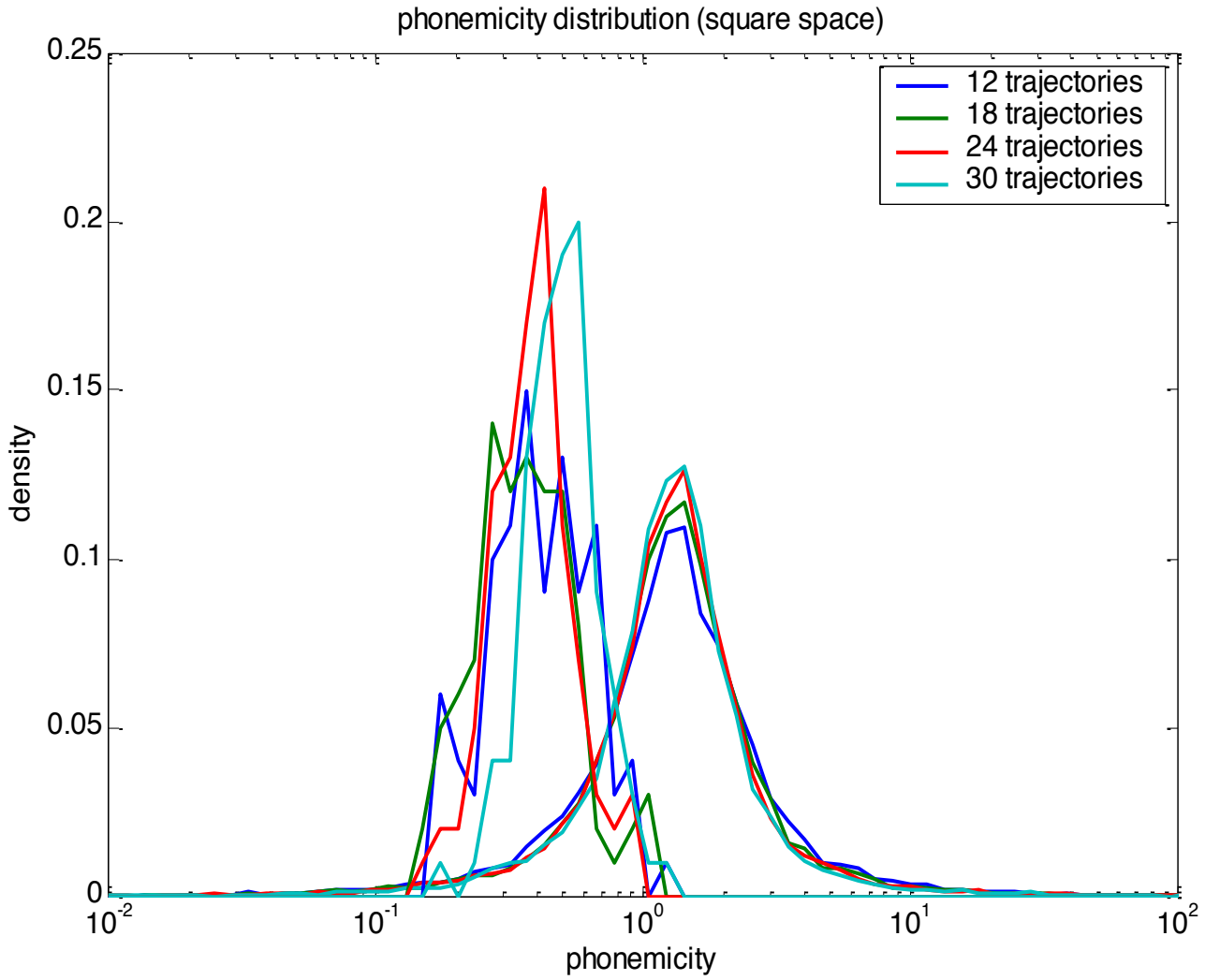
6



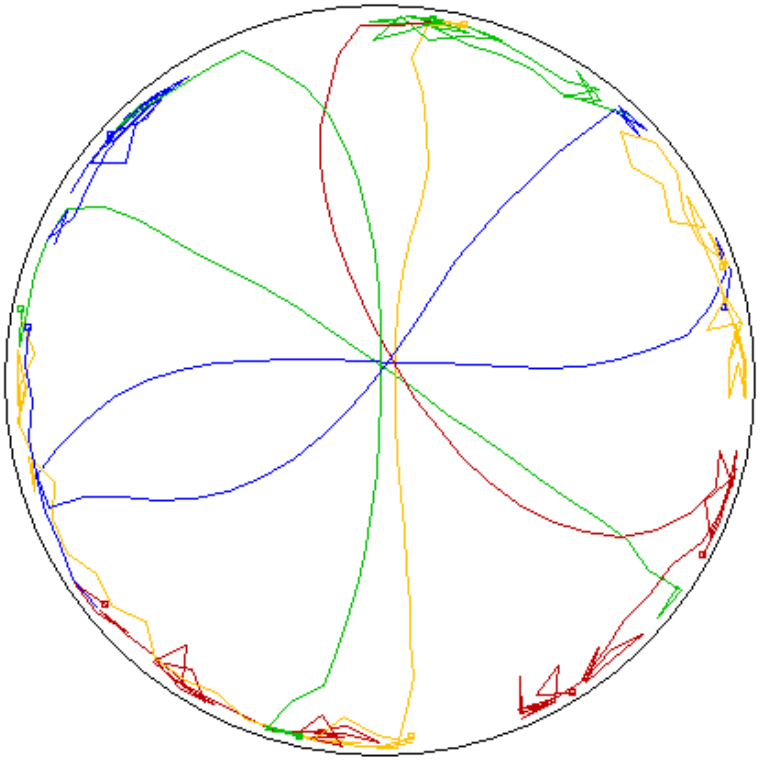
10

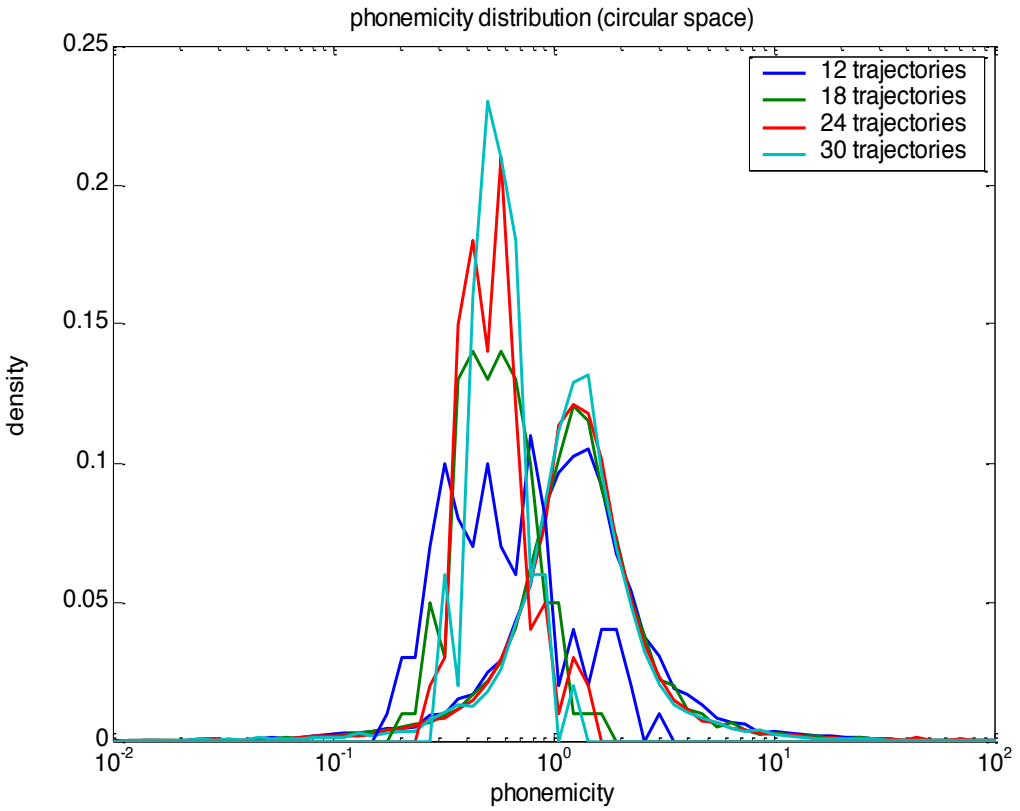


Random vs. optimal



Are corners necessary?



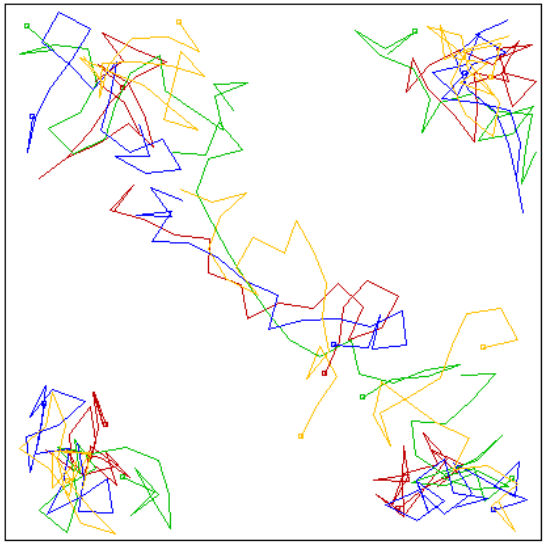


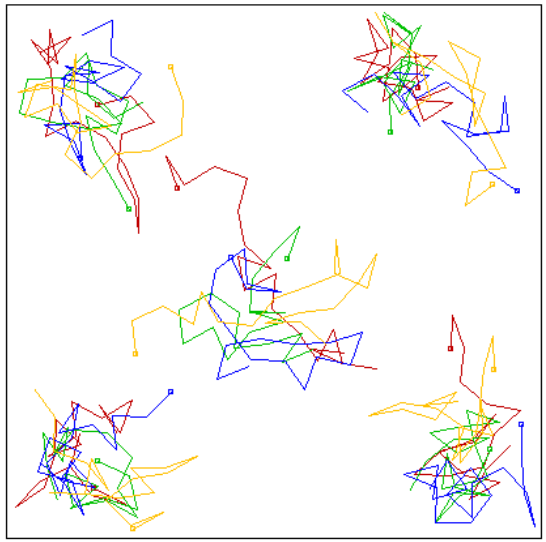
Zuidema & de Boer 2009, *J. Phon.* 37: 125

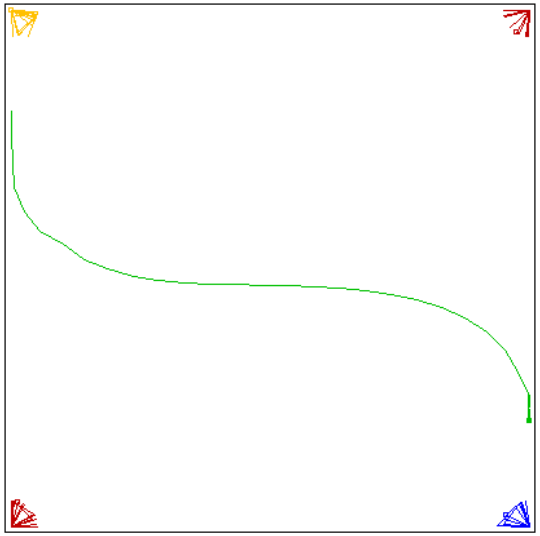
And in a population?

- Direct optimization is cheating in a sense
- What if self-organization in a population has to do the work?
  - Very calculation intensive

Optimisation and population







de Boer & Zuidema, 2010, *Adapt. Beh.* 18: 141-

The results…

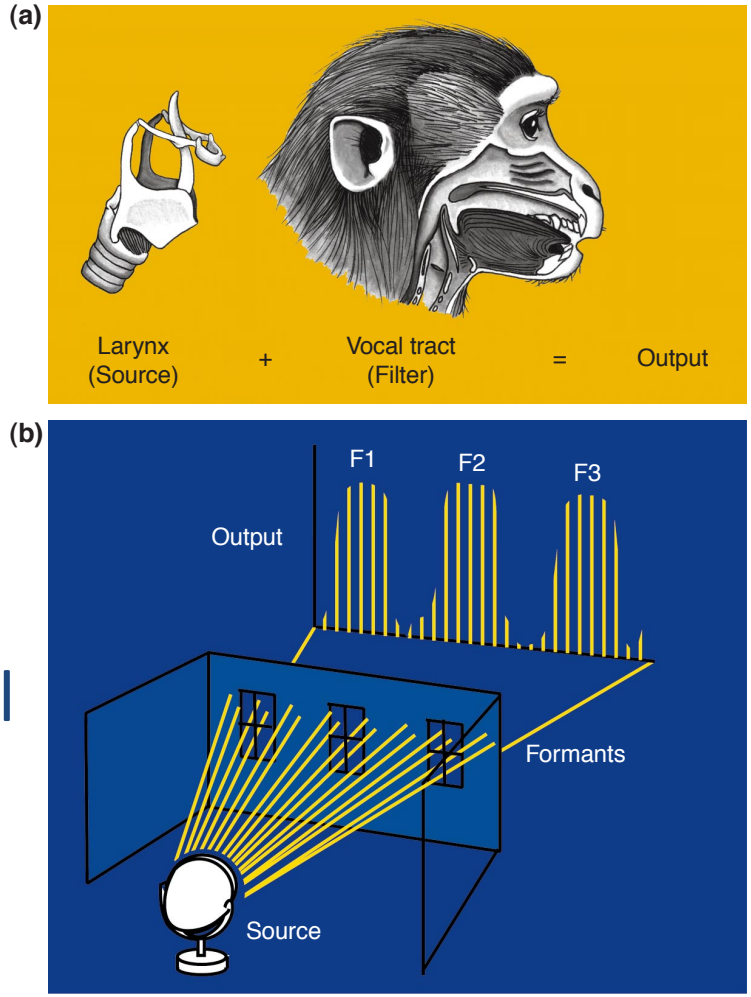
- Show that trajectories stretch out and corners are reused
- Looks like phonemic coding
- But the speakers are not aware of this

- Cultural precursor to biological adaptation to productive use of structure

3: An open problem

An open problem (1)

- The evolution of the vocal folds
  - We tend to take independence of vocal folds from the vocal tract for granted
  - But this appears to be special in humans



Fitch 2000

An open problem (2)

- Human vocal folds are very different from those of chimpanzees
  - But we really know very little about this
  - There is plenty of opportunity for observation, modeling and experiment

Kelemen 1969

Conclusion

- Modeling of evolution of anatomy can make a useful contribution to our understanding of language evolution
  - But perhaps not through the ultimate reconstruction of ancestral anatomy
  - Rather through exploration of what the effects of different adaptations could be
  - And through reconstructing sounds to investigate perceptual effects
- Virgin territory still exists
  - e.g. vocal fold evolution

Thank you