Formal and Computational Approaches to Phonology

Wednesday: Vowel Simulations

Julian Bradfield and James Kirby

University of Edinburgh

Evolution of Vowel Spaces

Quite early, people started trying to model the emergence of phonological categories.

Vowels are particularly popular (because they're much easier to model . . .).

Work from 70s onwards by Liljencrantz and Lindblom, Steels, and others. Large modern study around 2000 by de Boer's thesis. We'll use de Boer's general framework.

Questions of interest: why do languages so often produce vowel systems like those we saw on Monday (not Swedish!)? Is it a natural consequence of . . . of what?

How do we model such a thing?

And now for something completely different . . .

So far, we've been looking at 'pure', traditional, phonology: discrete, categorical, abstract systems.

Now we turn to the connexion of such phonology to real(?), continuous phonetics.

Vowels and the Vowel Space

What are phonological vowels? Do we believe in features? Or are they just different regions of vowel space? For now, the latter.

What is vowel space? How many dimensions? Articulatory or acoustic or perceptual?

Articulatory: traditional IPA chart classification: height, backness, rounding as three orthogonal components. What is the detail?

Acoustic: traditionally, identify vowels by *formants*. Approximately, log F_1 varies inversely with height, log F_2 varies inversely with backness, log F_3 varies inversely with rounding. *But it's much more complex than that*, and in ptic rounding affects F_1 and F_2 also.

Perceptual: how do we perceive vowels? Do we actually hear formants? (If so, how do we deal with variation?) Calling neuroscientists . . .

Vowels *per* de Boer

De Boer compromises between detailed physical modelling, and pure abstraction:

- articulation: is represented as three numbers in $[0, 1]$ giving height, backness and rounding;
- acoustic signal: is represented as four formant frequencies: these are calculated from articulatory positions using interpolations of real data;
- perception: is represented as F_1 and 'effective second formant', calculated from other formants (based on experimental data from approximations of vowels by two formants), on perceptual Bark scale;
- speakers map from perception to articulation by 'talking to themselves': trying to find articulations that produce the same percept.

It is interesting to observe that in his actual thesis (rather than the published book), he started with a much more detailed physical model. It didn't work ...

Steels' Imitation Game, *per* de Boer

We assume a population of individuals, who are evolving a common phonology.

Individuals have a discrete set of distinct vowel phonemes.

Individuals learn by repeated interactions with others. The pattern of an interaction is:

- Speaker S says a vowel v.
- Listener L hears sound of v, maps it to its own vowel v'.
- L says v' back to S.
- S hears sound of v', maps it to its own v''.
- S informs L 'out of band' whether $v = v''$.
-
    - if $v == v'$, L marks v' as more successful, and perhaps moves it towards v; otherwise
    - L marks v' as less successful, and creates a new vowel phoneme based on v.

Phonetics, Phonology and Learning

We also have to model the process of learning a phonology.

How does a child convert the heard sound into its own articulatory instructions?

How does it know when it's correctly making some distinction that the adult makes?

(And how does all this work given that child voices are very different from adult voices? Let's not go there.)

Noise, variation and tidiness

If transmission is perfect, nothing much will change. So we add noise to the signal:

- articulatory noise represents the inherent variability in articulation
- acoustic noise represents ... noise in the acoustic signal
- perceptual noise is arguably unnecessary

Gaussian noise would seem natural; however, de Boer uses uniform noise over $[-\psi/2, \psi/2]$ where the ψ are parameters of the simulation.

To encourage variation, speakers occasionally randomly add a new phoneme to their inventory (frequency by parameter).

To avoid proliferation of closely spaced vowels (Swedish!), speakers periodically tidy up by merging vowels that are perceptually close (by parameter).

Demo

Simplifying the model

Suppose we abandon the relatively realistic modelling of articulatory, acoustic and perceptual domains, and just say that vowels are points in the unit cube, with a perceptual distance metric which squashes (a) backness when low (b) rounding compared to other dimensions (like the IPA vowel cuboid).

How much changes?

Demo

- 20 agents for 10000 interactions, parameters set to merge articulatory nearby vowels (in a cube). Run.
- The same, but vowels merged in perceptual space (vowel chart). Run.
- The same, with stronger mutual accommodation between speakers. Run.

Now what have we learned?

Discussion

If the demo went well, we saw vaguely natural-looking vowel systems emerging.

- What have we learned from doing this?
- Did we have a falsifiable hypothesis? What was it?
- How does the success (or otherwise) of our ‘experiment’ depend on
  - the architecture of the model
  - the parameters within the decided architecture
- Whatever we’ve learned, how do we transfer it back to the real world?
- Compare and contrast with, e.g., an astrophysics simulation of galaxy formation; a predator–prey simulation in population dynamics; an economic forecasting model; etc. etc.

Getting a bit more phonological . . .

An interesting use of simulations is to try to support the psychological reality of phonological concepts.

Boersma and Chládková 2010

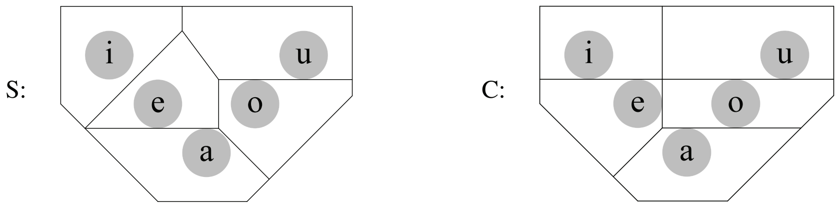
Simulation framework is agents learning a 5-vowel system via an OT phonological grammar in Boersma's interconnecting module version.

- Learners learning points in vowel space have 'diagonal' perceptual boundaries between vowels.
- Learners learning categorical features (high/back etc) have horiz./vert boundaries.
- In reality, the latter happens (Savela 2009).
- They suggest this is evidence for features.

Moreover . . .

B&C on Spanish and Czech

- Spanish and Czech both have classic 5-vowel systems.
- But phonology suggests Czech /a/, /e/ are [back] and [front], but Spanish /a/, /e/ are [central]
- which (per previous) should affect perceptual boundaries.



- This is what the simulation does with Czech and Spanish featurally specified targets.
- It's also what B&C find in real speakers!

But is there a phonetic explanation?

We set up a simulation using learning via imitation game again, but:

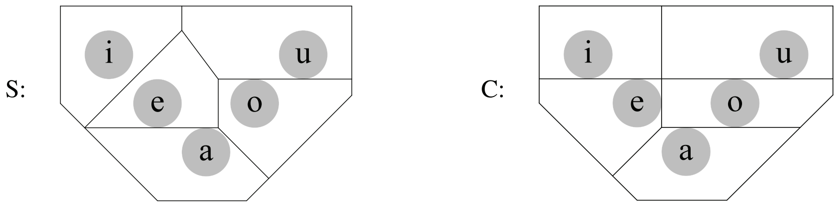
- We distinguish children from adults (don't learn) and have a dynamic population.
- The agents have a richer notion of vowel: articulatory prototype, and perceptual regions (convex polygons extended as they hear new exemplars).
- But the vowels are still simple and abstract (no phonetic detail, just F_1 and F_2.

We seed the initial adult population with Czech or Spanish articulatory prototypes, and ask:
Is it stable? What are the perceptual boundaries do the agents develop?

Four simulations

All specified by initial articulatory prototypes:

- A pure 5-vowel system Run.
- with slightly raised e,o Run.
- A Spanish 5-vowel system Run.
- A Czech 5-vowel system Run.



appropriate different perceptual boundaries can arise as purely emergent phonetic consequences of vowel positions – no features in sight!