Acquisition and Complexity of Phonemes and Inventories

Julian Bradfield

Introduction

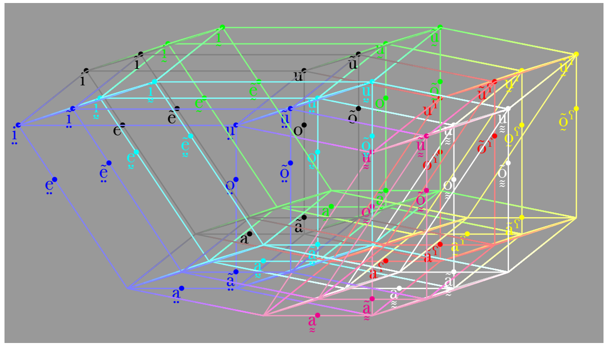
The first question in a computational simulation of phoneme acquisition is, what is a phoneme, and what structure is placed on the phoneme space. It might be as simple as a discrete set. Or, as in de Boer's (2001) well known work on the evolution of vowel systems, it might be a two or three dimensional continuous space, with a metric, within which phonemes are points, clusters, or distributions. For consonants, it is less obvious what to do, as the mapping between articulatory, acoustic and perceptual space is complex and far from one–one. As there is some evidence that infants recognize and generalize some features better than others, one might place consonant phonemes in feature space. On the other hand, the ability to generalize over features appears to degrade, perhaps contributing to the difficulty of acquiring new phonemes later in life, which suggests that at some point the mental phoneme space becomes less structured.

The structure of phoneme space becomes especially interesting for languages with highly complex inventories, such as the Caucasian languages and the San languages – particularly since those inventories are usually rather neatly structured when laid out in IPA style charts. The San language !Xóõ has perhaps the world's most complex vowel inventory. According to Traill (1985), the phonetic vowel space comprises a five-vowel system combined (with some restrictions) with any combination of three distinctive voice qualities and also nasalization. Some simplification occurs on a plausible classical phonemic analysis, but there still remain 37 distinct vocalic phonemes, some of which are so rare that they may well not be encountered for some years, if at all. A similar problem arises with the better known click inventory.

I recently (Bradfield 2009) suggested that even in adulthood, the !Xóõ vowel inventory is better understood as being structured into concurrent 'phonemes', with each 'voice quality' being a 'phoneme' (in the style of autosegmental phonology, but less drastic). I argued, on only intuitive grounds, that it would ease the otherwise challenging acquisition problem. This presentation presents a first attempt to produce some quantitative information to check this idea.

The !Xóõ vowel space

According to well supported arguments by Traill (1985), the !Xóõ vowel space comprises five basic vowels /a e i o u/ which usually occur in pairs in content-word stems (either in CVCV or CVV form). In addition, stems may carry, primarily on the first vowel, any combination of breathiness, pharyngealization and laryngealization. (The famous 'strident' or epiglottalized vowels are interpreted as phonological breathy pharyngealized vowels.) On the second vowel, there may be nasalization, which spreads to the first. Phonetically, then, the space has a six-dimensional structure (ignoring rounding):



Simulation framework

We work in the style of de Boer, with a single 'adult' speaker (representing a population of similar adults), and 'learners' trying to acquire the adult's inventory. De Boer, following Steels (1997) and others, uses 'imitation games' – the learners try to match a vowel said by the adult, and receive extra-linguistic feedback on whether they succeeded; if not, they adjust their phoneme inventory by shifting a vowel or adding a new vowel, depending on various criteria.

Experience with de Boer's work suggests that his relatively detailed modelling of articulation and acoustics is not crucial, and moreover we are concerned only with the single issue of the topology of the phoneme space. We therefore simplified the model by mapping articulatory space directly into acoustic and perceptual space. Simply to reduce computation times, we also ignored nasalization, concentrating on the first-vowel qualities.

The models

We set up three models:

1. a simple 5-vowel system (reduced to height and backness, ignoring rounding, again for simplicity)
2. a 40-vowel system comprising 5 vowels together with any combination of 3 voice qualities, where the learners do not structure the phoneme space (i.e. they are simply learning 40 vowels in a 5-D space), which corresponds to a simplified version of Traill (1985)'s phonemic analysis
3. the same number of phonetic vowels, but with learners who identify voice qualities independently of vowel quality. That is, they 'hear', and have a mental representation of, each voice quality separately. This corresponds to the mental model proposed in Bradfield (2009).

Measurements

We measured the average number of interactions required for a learner to acquire fully the adult inventory. A learner was deemed to have acquired the adult inventory at the point where it could engage in 500 consecutive interactions without failure. (We could also use a deterministic definition by matching inventories, but for ease of implementation we have not yet done this.) The results here are from runs with 100 learners. (Learners do not interact with one another, so this is equivalent to 100 single learner runs.)

Results

The mean number of interactions required to achieve adulthood (i.e. the number of interactions before the 500 successful ones), and standard deviations, were for each model (to 2 s.f.):

1. (simple 5-vowel model) mean 41, s.d. 18.
2. (unstructured 40-vowel model) mean 730, s.d. 180.
3. (structured 40-vowel model) mean 150, s.d. 67.

The distribution is of course Poisson-like rather than normal; the high s.d.s reflect a long tail above the mean.

Conclusion

The results provide significant support for the suggestion that a structured inventory is much easier to learn. Interestingly, the agents take nearly twenty times as long to learn the 40-vowel model (2) as the base 5-vowel model (1). But the structured 40-vowel model (3) takes only four times as long as the 5-vowel model to learn.

Future work includes refinement of the model, investigations of its robustness, and extension to consonants and other inventories.

References

Bradfield, J. (2009). Clicks, concurrency and the complexity of Khoisan. Presentation at 17th Manchester Phonology Meeting.

de Boer, B. (2001). *The Origins of Vowel Systems.* Oxford University Press.

Steels, L. (1997). The synthetic modelling of language origins. *Evolution of Communication 1(1)*, 1–35.

Traill, A. (1985). *Phonetic and Phonological Studies of !Xóõ Bushman.* Hamburg: Buske.

School of informatics