

## Principal components of sound systems

Category: either oral or poster

Phoneme inventories of the world's languages as depicted by the UPSID database (Maddieson 1984, Maddieson and Precoda 1990) are analyzed using multivariate statistical techniques of principal components analysis and k-means and hierarchical clustering. The first two meaningful principal components, representing dimensions that account for the most variance in sound systems but are not caused by differences in typological frequencies of phonemes, are found to separate languages into three large clusters, distinguished by glottal articulations present in the stop inventory (with secondary and primary glottal articulations patterning similarly) and the sonority of other types of sounds present in the language. Other dimensions that are shown to separate languages into distinct clusters are 1) presence/absence of prenasalized voiced stops, 2) presence/absence of nasal vowels, and 3) presence/absence of schwa. Interestingly, obstruent inventories separate into distinct clusters as do vowel inventories but inventories of sonorant consonants do not, suggesting that the goodness of sonorant consonant inventories is more evenly distributed across the space of possible sonorant consonant inventories whereas spaces of vowel systems and consonant systems host several distinct attractor locations that are more optimal than other locations in the space.

Clustering analyses, which automatically categorize sound systems and phonemes, are shown to reveal both areal groupings of languages, for instance, categorizing together genetically unrelated languages of India. Groupings of phonemes are often interpretable as the conjunction of featural descriptions and typological frequency, especially when clustering analyses are conducted within phoneme categories defined by manner of articulation / sonority. For instance, cardinal vowels /i/, /a/, and /u/ are clustered together, with /o/ and /e/ forming a second subcluster. This clustering is also found for nasalized vowels, long vowels, fricated vowels, and laryngealized vowels, each complication defining a cluster at a higher level.

### References:

Maddieson, I. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, I., and K. Precoda 1990. Updating UPSID. *UCLA Working Papers in Phonetics* 74: 104-114.