Modeling Distinctive Feature Emergence

Jeff Mielke
University of Arizona

1. Natural classes

It is well known that classes of sounds which take part in sound patterns tend to be "natural" in some way. For example, phonetically coherent classes like /m n ŋ/ and /u o ɔ/ seem to recur in different languages, while more arbitrary groupings like /m n tʃ/ and /ɨ ʔ k^w/ are less common. Different explanations for this observation have been offered, as in (1-2):

(1) **An innatist claim:** Common classes (or common sounds patterns) can be described using a conjunction of distinctive features (e.g., McCarthy 1994, Clements and Hume 1995).

(2) **An emergentist claim:** Common classes (or common sounds patterns) result from common historical changes (e.g. Blevins 2004, Mielke 2004).

While these two claims are not contradictory, this paper provides an account of the abundance of natural classes without recourse to innate features. Evidence is provided in Section 3 from a simulation of class emergence which is based on readily-observable aspects of phonetic similarity, provided in turn by a phonetic similarity metric detailed in Section 2.

One way to examine the claim that common classes are the ones which are statable in terms of distinctive features is to conduct a survey of classes that are involved in sound patterns, and see how many of these can be accounted for in terms of features in different theories.[1] Mielke (2004) reports on *phonologically active classes* from grammars of 561 languages, about 17,000 sound patterns. A phonologically active class is defined as any group of sounds which, to the exclusion of all other sounds in a given inventory, (a) undergo a phonological pattern, or (b) trigger a phonological pattern. The database contains 6077 distinct classes fitting this description. All of these classes are naturally occurring, and the terms "natural" and "unnatural" will only be used in reference to a specific feature theory. A class of sounds is *natural* with respect to a particular theory if the class is statable as a conjunction of features in that theory. A class of sounds is *unnatural* with respect to a particular theory if it is not statable as a conjunction of features, but rather requires special treatment, such as disjunction or subtraction of natural classes, or is unstatable.

The results show that unnatural classes are not particularly rare. 3640 classes (59.90%) are natural according to the feature system of *Preliminaries to Speech Analysis* (Jakobson, Fant, and Halle 1954), while 4313 classes (70.97%) are natural according to the feature system of *The Sound Pattern of English* (Chomsky and Halle 1968), and 3872 classes (63.72%) are natural according to Unified Feature Theory (Clements and Hume 1995). 1496 classes (24.65%) are unnatural according to *all three* of these feature theories.

Figure 1 shows the distribution of natural and unnatural classes in terms of the Unified Feature Theory feature system, which in this case is fairly representative of the three systems examined. In this chart, unique feature specifications are arranged along the x-axis in order of decreasing frequency. Light bars represent classes which are natural in Unified Feature Theory, and dark bars represent

* Thanks to Diana Archangeli, Adam Baker, Chris Brew, Robin Dodsworth, Mike Hammond, Jay Myung, Natasha Warner, and Andy Wedel for helping with this. This research was supported in part by James S. McDonnell Foundation grant #220020045 BBMB to Diana Archangeli.

1 There are also methodological reasons for natural classes to appear more common. For example, natural classes are easier to count than unnatural classes. /m n/, /m n ŋ/, and /m n̪ n ɳ ɲ ŋ/ are all counted as nasals, but it is less obvious what type of class /m n tʃ/ is an instance of, and what another example of this class would be. Further, natural classes are usually more rewarding to write phonology papers about.

 Proceedings of the 24th West Coast Conference on Formal Linguistics, ed. John Alderete et al., 281-289. Somerville, MA: Cascadilla Proceedings Project.

classes which are unnatural in the theory, i.e., those whose specification requires disjunction or subtraction of feature bundles. The height of each bar indicates the number of occurrences in the database of classes corresponding to that feature specification. Both axes are on a logarithmic scale, so while the width of each bar indicates the number of unique feature specifications sharing the same frequency, bars on the left side of the chart appear inherently wider, but the number of feature specifications represented by each bar is apparent from the x-axis scale.

If recurrent natural classes are simply the classes that can be represented with a conjunction of innate features, there should be a dropoff between common natural classes and uncommon unnatural classes. It is clear from Figure 1 that no sudden dropoff is found. There is not even a boundary between classes which the theory finds to be natural or unnatural. The two are fairly well interleaved, with several unnatural classes being more common than most natural classes, and the vast majority of logically possible natural classes unattested. In other words, most of the feature bundles which are possible in these theories have a frequency of zero and are therefore are not included in Figure 1. The fact that logically possible classes are unattested is not remarkable in itself, but it is interesting in light of the fact that many classes predicted not to occur are attested many times over.

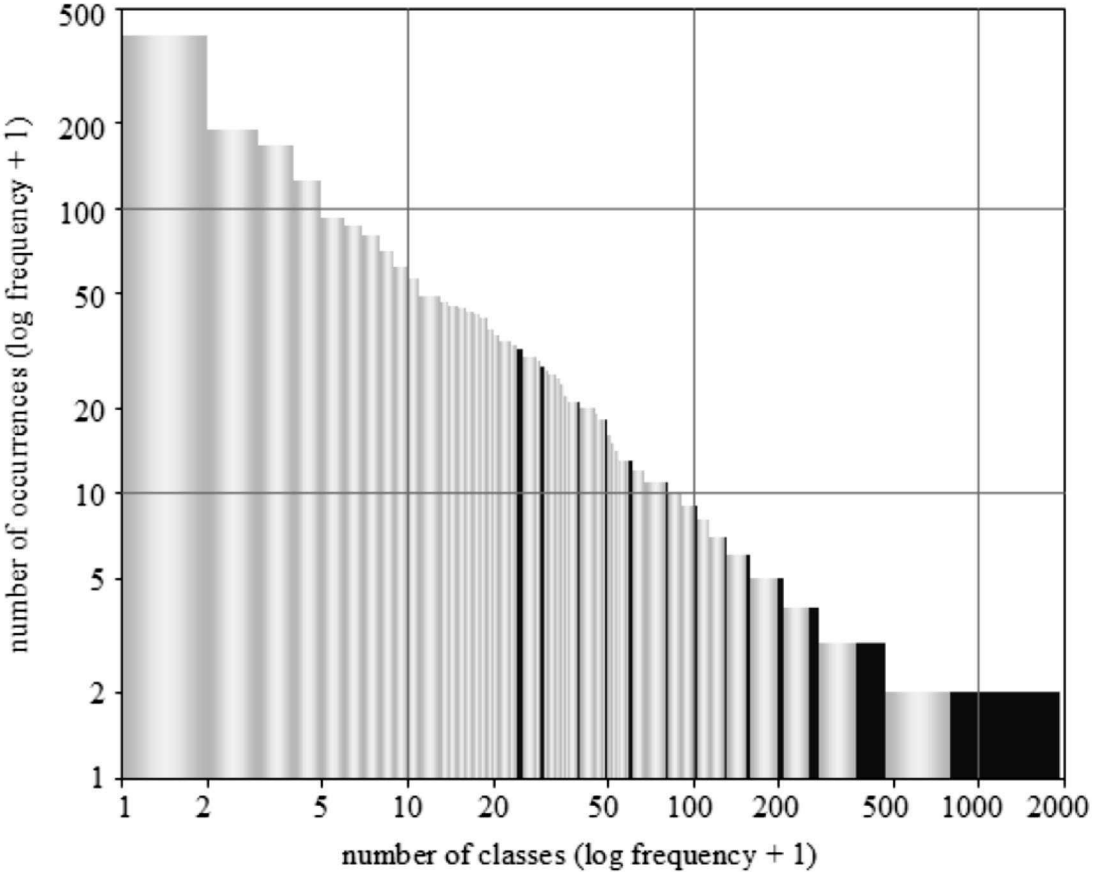

Figure 1. Survey results: no boundary between natural and unnatural

There are reasons to expect natural classes to be frequent anyway, with or without innate features. Two opportunities for naturalness to be favored are (1) the inception of a sound change and (2) a change in an existing sound pattern.

Opportunity #1 is the inception of a sound change. Phonological patterns can result from a change characterized as phonologization (Hyman 1977), conventionalization, or exaggeration (Janda 1999) of a previously "insignificant" phonetic effect. Phonetic effects involve physiology, aerodynamics, etc.. Most humans have very similar vocal tracts and auditory systems and all are subject to the same laws of physics. Likewise, languages have similar classes involved in similar phonetically grounded sound patterns which reflect these facts (see also e.g. Dolbey and Hansson

1999). For example, physiology and inertia dictate that coarticulatory vowel nasalization is likely to be caused by adjacent nasal segments which are produced with velum lowering. Consequently, nasality alternations will tend to involve the familiar class of nasals, regardless of what feature system is used in the synchronic grammar.

Opportunity #2 is a change in an existing pattern. Classes involved in sound patterns may change over time, and one way for this to happen is for learners to arrive at the "wrong" generalization about a sound pattern, and for the "wrong" generalization to become the prevailing version for a speech community. When this happens, the "wrong" generalization has become "right". The result may be a phonetically natural class that is not necessarily related to the original phonetic basis for the sound pattern. For instance, Mielke (2004) argues that the class of depressor consonants in Zina Kotoko (which surprisingly includes implosives) (Odden 2002) is a result of the mislearning of a phonetically natural tone lowering pattern originally conditioned only by plain voiced obstruents and sonorants. The resulting class includes voiced implosives, which, while forming a phonetically coherent class with other voiced consonants, could not have been part of the original phonetic basis for tone lowering, because implosives cause phonetic F0 *raising*, not lowering.

What is significant about forming new generalizations is that all generalizations are not equally plausible. There are bizarre classes, but mistakes are expected to favor classes of sounds that have something in common, e.g., /m ŋ/ can plausibly be mislearned as /m n ŋ/, but /m ŋ ʧ/ seems to be a less likely mistake. This means that classes of phonetically similar segments should appear with greater than chance frequency, and arbitrary groupings of sounds likely have no reason to be favored. This generalization bias, along with the well-documented effects of phonologization, predict that phonetically natural classes should be common relative to phonetically unnatural classes, regardless of whether there are innate features in Universal Grammar. It will be shown in the following sections that this bias toward phonetically natural generalizations can be modeled without stipulating features. But this cannot be done without first defining what it means to be phonetically similar.

2. Phonetic similarity

Defining phonetic similarity is a prerequisite for using it to account for phonological observations. Frisch, Broe, and Pierrehumbert (1997) offer a phonetic similarity metric, one that is based on the number of shared natural classes, and consequently requires predetermined features. It is therefore not applicable to the question of whether classes emerge without innate features. Instead, what is needed is a similarity metric based on objective measurements of sounds. In order to develop such a metric, it is necessary to choose sounds to measure and to choose ways to measure them.

There are 906 distinct symbol/diacritic combinations used to transcribe segments in the class database. It is impractical to include all of these in a first attempt, so it is necessary to try to minimize the number of segments while maximizing the number of languages whose entire inventories are included in the set of segments which are chosen. 416 segments (46%) occur in only one language each, and can readily be excluded. In order to choose a minimal set of segments which will be maximally useful for making inferences about human languages, it is helpful to identify the segments which are found in the segment inventories of the most languages. Local maxima of the segment/language ratio were discovered by starting with 610 language varieties (those represented in the class database) and 906 segments from the survey in Section 1 and iteratively removing segments which occur in only one language, and then removing languages whose inventories are no longer subsets of the remaining segments. The local maximum of 63 segments (Table 1) was perceived to be small enough to be practical and large enough to be useful. This inventory of 63 segments (7.0% of the segments in the survey) manages to cover 124 of the survey's language varieties (20.3%).

In choosing measurements, it is ideal to get a "360-degree" view with perceptual, acoustic, and articulatory measurements of these 63 segments. The perceptual component of the model is saved for the future because 63 segments amount to 1953 pairs of different segments, which is a lot for a subject to listen to. Further, adult subjects typically have phoneme categories from their native languages. Infant subjects would provide a way around this problem, but exacerbate the problem of having so many pairs to listen to. The present model is based on raw acoustic similarity, and articulatory measures of tongue and lip position and oral and nasal airflow.

p	b			t̪	d̪	t	d	ʈ	ɖ	c	ɟ	k	g	kʷ	ʔ
	ɓ					tʼ	ɗ					kʼ			
						ts	dz			tʃ	dʒ				
ɸ	β	f	v			s	z			ʃ	ʒ	x			h
	m				n̪		n		ɳ		ɲ		ŋ		
						r	ɾ		ɽ						
									ɻ		j			w	
					l̪		l		ɭ		ʎ				
										i	iː	ɨ		u	uː
										e	eː			o	oː
										ɛ		ə		ɔ	
										æ		a	aː		

Table 1. A superset of 124 inventories

The 63 sounds were produced by a trained linguist in three different segmental contexts: a__a, u__u, and i__i, resulting in 189 tokens, which were subjected to acoustic and articulatory analysis. Audio, video, and ultrasound recordings were made simultaneously, using an Audio-Technica PRO 49Q condenser microphone through one channel of a Symetrix 302 dual microphone preamplifier, a 1 Megapixel Sony MiniDV Digital Handycam, and a SonoSite TITAN ultrasound unit with a C-11/7-4 11-mm broadband curved array transducer. These channels were combined using a Videonics MXProDV digital video mixer and recorded on a Sony MiniDV Digital Video Cassette Recorder. Individual video frames were selected and the audio channel was extracted as 16-bit 44.1kHz mono .wav files using Final Cut Express 2. In a second session with the same speaker, nasal and oral airflow were measured for productions of the same stimuli, using oral and nasal masks and SCICON R&D's Macquirer 516 transducer interface system and Macquirer software.

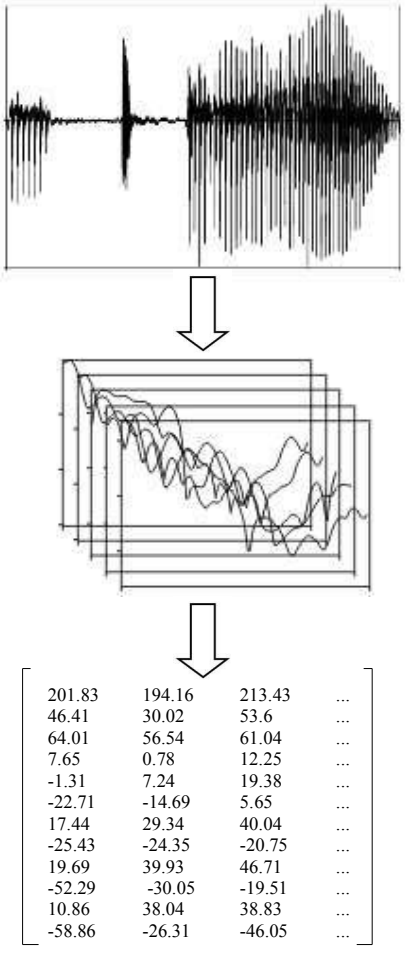

Figure 2. Converting waveforms to spectra and then matrices.

To compute acoustic similarity, waveforms were first converted into matrices (Figure 2). The 189 waveforms were chopped into overlapping 15 ms slices, 111-239 slices per waveform, using Praat. Spectra of these slices were run through Mel filters (spaced 100Hz apart), and converted to 12 Mel-scaled cepstral coefficients (again using Praat). Thus, 189 waveforms were reduced to 189 matrices ranging in size from 12×111 to 12×239. The Mel filtering introduces an auditory scale which is more similar to how these waveforms are perceived by mammalian listeners. Next, the set of matrices was converted into a set of distances. Distances were calculated between pairs of matrices using a Dynamic Time Warping (DTW) algorithm (see e.g. Holmes and Holmes 2001) implemented in Python, as shown in Figure 3. Due to DTW, spectrally-similar intervals are matched, even if the durations are not the same. The end result is three 63×63 distance matrices, i.e. the set of segments crossed with itself once for each segmental context. These acoustic distances were then converted to a small number of acoustic dimensions using SPSS's multidimensional scaling (MDS) algorithm. The best model has three dimensions (i.e., little improvement in S-stress or R^2 is achieved by adding more

dimensions). The dimensions extracted from the waveforms can be interpreted roughly as sonorous/mellow vs. sibilant, grave vs. acute, and low formant density vs. high formant density.

Articulatory similarity was computed on the basis of ultrasound images of the tongue and palate, video images of the face, and oral and nasal airflow measurements. Combined face/ ultrasound video frames were extracted from the middle of each target segment as .jpgs using the QT Extractor plugin (by Adam Baker) for ImageJ (by Wayne Rasband). In these images, the speaker's palate was superimposed on each image (using Palatron (Mielke *et al.* 2005), as shown in Figure 4. The point at which the tongue and palate surfaces are closest was identified, allowing measurement of the location of the constriction on the palate and the distance between the tongue and palate at this point. The height of the opening between the lips was measured, and the ratio of nasal airflow to oral airflow was computed for each token.

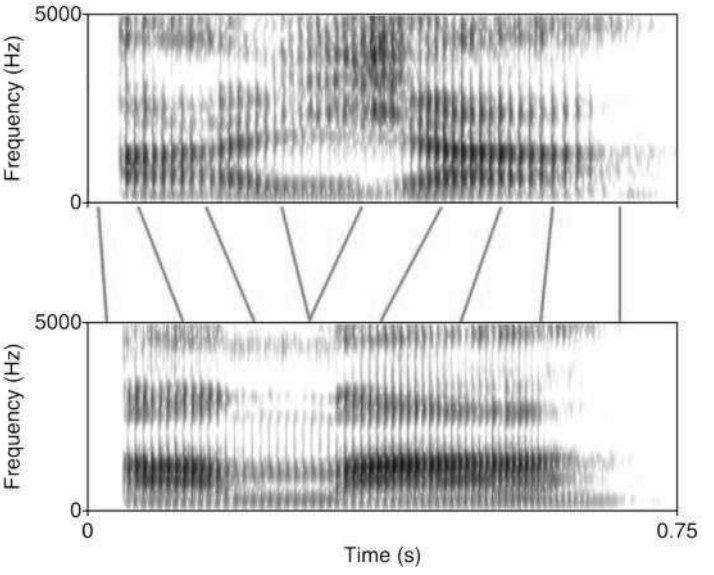

Figure 3. Dynamic time warping.

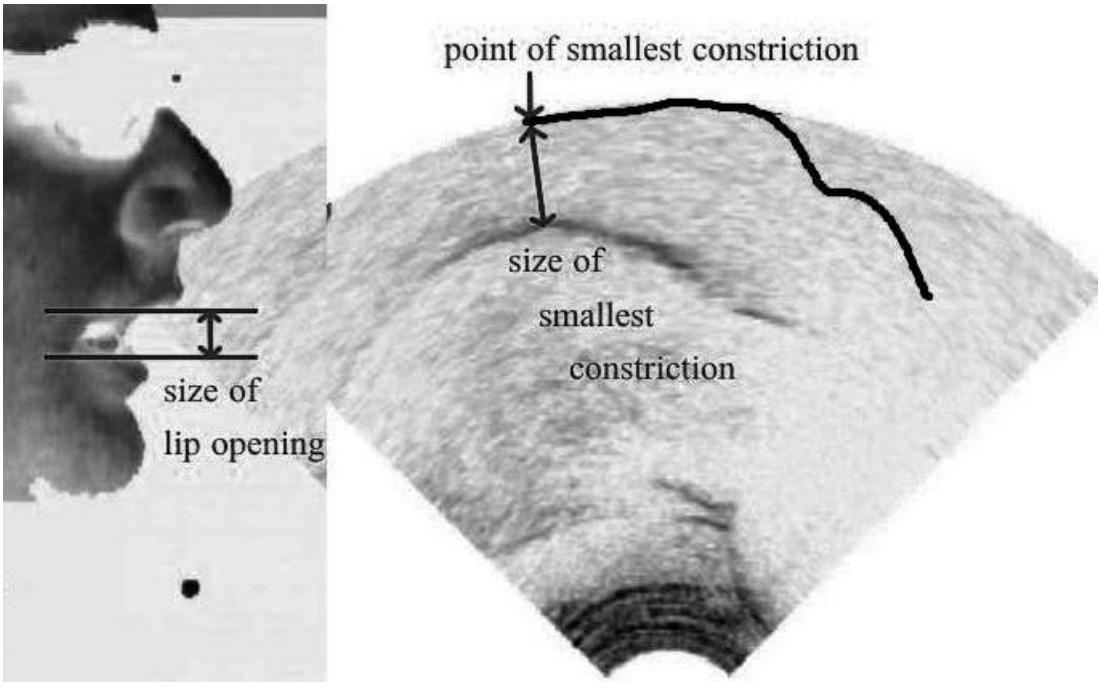

Figure 4. Tongue and lip measurements.

The ranges for each of the seven measurements were normalized so that each would cover a range of about 5 arbitrary units and be centered on zero. Based on these acoustic and articulatory measurements, each segment is represented by a 7-dimensional vector which contains three acoustic dimensions (1. sibilant vs. sonorous/mellow, 2. grave vs. acute, and 3. high vs. low formant density) and four articulatory dimensions (4. oral constriction location, 5. oral constriction size, 6. lip

constriction size, and 7. nasal/oral airflow ratio). The phonetic distance between any two segments can now be measured as the Euclidean distance between those two segments' vectors. For example, it is possible to state that the distance from [a] to [æ] (1.30) is less than distance from [a] to [e] (2.13). Similarly, the distance from [l] to other coronal sonorants such as [ɻ] (1.55) and [n] (1.57) is less than the distance from [l] to other segments such as [t] (1.92), [k] (2.34), and [a] (2.13). Clearly [k] and [a] are not different from [l] in the same way, and different types of similarity can be quantified by measuring distance in various subsets of the dimensions.

Two-dimensional projections of the model (created by multidimensional scaling, so that they can easily be displayed on paper) are shown in Figure 5. When projected in two dimensions, the short vowels resemble a familiar acoustic or articulatory vowel space (Figure 5a). The consonants group fairly well into manner classes (Figure 5b), with similar classes (such as all of the sonorant consonants) clustering together, and with the rhotics ranging visibly from the most obstruent-like to the most canonically sonorant. The two-dimensional projection of the 7-dimensional model does not differentiate places of articulation other than coronal very well (Figure 5c), but when limited to the four articulatory dimensions (Figure 5d), a 2-dimensional projection shows labial, coronal, and velar consonant breaking into separate corners, with glottals and [ɻ] in the middle.

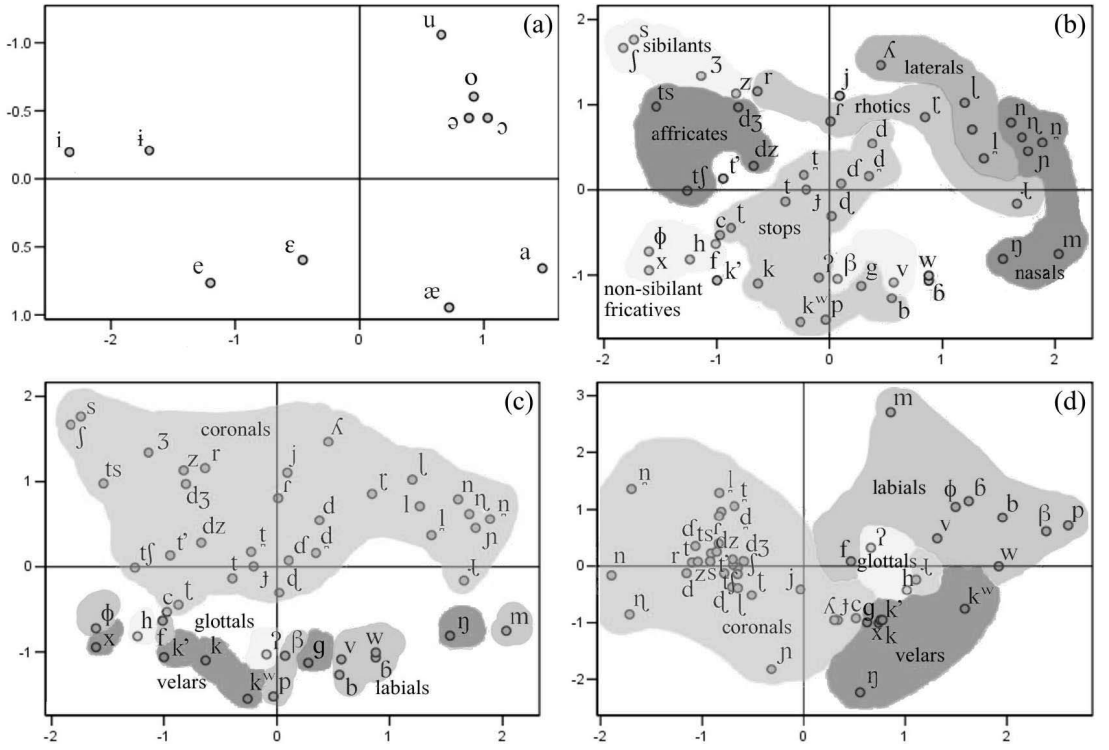

Figure 5. (a) Short vowel similarity in two dimensions; (b) consonant similarity in two dimensions shaded by manner; (c) consonant similarity in two dimensions shaded by place; (d) articulatory similarity in two dimensions shaded by place.

3. Modeling class emergence

The similarity metric is used to model the emergence of natural classes. The simulation consists of an inventory of segments, a lexicon, and a class of sounds which participate in an arbitrary sound pattern, and a learner, implemented in Python, who can tell how similar sounds are. The learner's task is to figure out what segments are in the class. Given a class of sounds such as /p t k/ selected from an inventory such as (3a), the learner tries to infer the class by observing the phonological behavior of words randomly selected from the lexicon as in (3b).

(3) a.

p	t	k	ʔ	i		u
	s			e		o
m	n				a	
	r					
	j	w				

b.

iui	/u/ doesn't participate
usu	/s/ doesn't participate
apa	/p/ **participates**
unu	/n/ doesn't participate
uʔu	/ʔ/ doesn't participate
utu	/t/ **participates**
	etc.

The lexicon is composed of a target segment drawn from the inventory preceded and followed by [a], [u], or [i]. Each word is represented by a 7-dimensional vector which is a combination of values from its target segment and from the specific *token* composed of the target segment and vowel context. This allows some coarticulatory effects on the target segment caused by the flanking vowels to be represented in the data. The acoustic and airflow dimensions (which are expected to be similar across a segment's three tokens) are averaged across the target segment's three tokens, but the other articulatory dimensions appear as measured for each token. As a result, phonetic properties which are consistent across tokens of the same segment (such as oral constriction location in coronal consonants) are consistently reinforced, while phonetic properties that are dependent on context (such as oral constriction location in [h] and labial consonants) generally cancel each other out, rather than reinforce a meaningless average tongue position for these segments.

The learner generates a hypothesis based on the words it is exposed to, and may correctly infer the class on the basis of incomplete data. This is possible because the learner does not assume that an observation is relevant only to the particular token or segment it observes, and applies knowledge of a segment's behavior to others which are phonetically similar to it. Each observation of a segment's behavior generates a normal distribution along each phonetic dimension. The level of activation of each segment in the inventory depends on its proximity to the observed segment in each dimension. The activation of a segment is the sum of its level of activation in each of the seven dimensions. Within a dimension, the activation of a segment is a function of the distance between it and the target segment in that dimension, as in (4). This is illustrated in Figure 6, showing the effects of an observation involving /p/ on its neighbors along the "acute"/"grave" dimension.

(4) $$activation = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(distance^2)/(2\sigma^2)}$$

Some adjustable parameters are the width (σ) of the activation function and the value of negative evidence (the fraction of the activation value that is subtracted when a segment fails to participate in the sound pattern). Decreasing width reduces the effect segments have on their neighbors, and increasing the value of negative evidence reduces the likelihood that positive evidence from neighbors will overwhelm direct evidence that a segment does not participate. Both of these discourage the learner from overgeneralizing. Additionally, sounds can have different frequencies in the lexicon (low frequency means that much of the learner's knowledge of a particular sound is determined by indirect evidence), and the learner's output can be used as part of the learner's input, allowing the "wrong" generalization to become the "right" generalization.

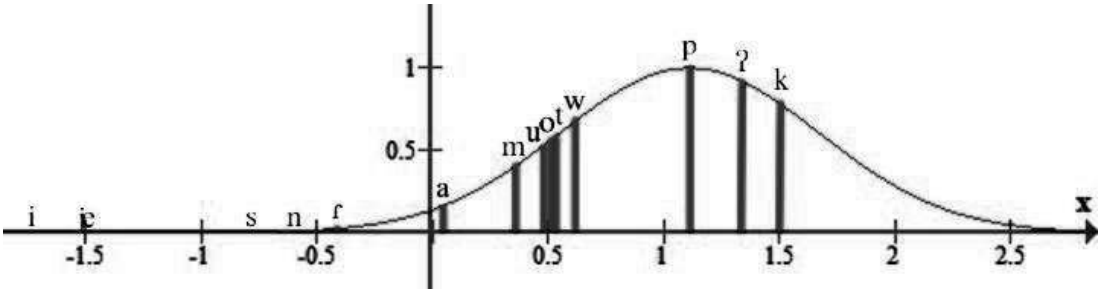

Figure 6. Reacting to data.

Similarity can be measured with respect to different subsets of dimensions, and in most cases some dimensions are better than others for distinguishing apparent participants from apparent non-participants. The learner capitalizes on this by comparing the means and the variance of what it understands to be the classes of participating and non-participating segments. A good dimension is one for which the means of the apparent participants and apparent non-participants are far apart, and the variance within each set is small. The learner responds to a good dimension by focusing more attention on it (its activation function gets taller and narrower). Unhelpful dimensions see their activation functions get flatter and wider, and if they continue to be unhelpful, they are eventually ignored almost completely.

To test whether the model can favor natural classes, the learner is given natural and unnatural initial classes, and the larger inventory in (5). The parameters are held constant, and the results are checked after 500 iterations. This is repeated 5 times each for a variety of classes.

(5)

p		t	tʃ	k		i		u
b		d	dʒ	g		e		o
	f	s	ʃ		h	ɛ		ɔ
	v	z	ʒ			æ	a	
m		n		ŋ				
		ɹ	j	w				
		l						

Table 2(a) shows a series of trials in which the parameters are set to encourage overgeneralization, and the learner is given the class /m ŋ/ as an input in a language which also has /n/. The /n/ gap tends to be filled, creating the phonetically natural class of nasals. When the class of nasals is given as the initial class with the same settings, it tends to remain. Both cases tend to produce natural classes. Table 2(b) shows how the /n/ gap persists under parameter settings which discourage generalization. This situation is typical of acquisition, where learners usually acquire a grammar which accurately produces the observed data rather than impose a new generalization on it. But what is interesting is what happens in the rare case when a new generalization does take hold. In Table 2(a), mislearning creates the natural class of nasals.

(a) σ = 0.5; neg. ev. = -0.07 (“Generalize”)		
Input:	m, ŋ	m, n, ŋ
Output:	m, n, ŋ	m, n, ŋ
	m, n, ŋ	m, n, ŋ
	m, n, ŋ	m, n, ŋ
	m, n, ŋ	m, n, ŋ, b, p, u, w,
	m, n, ŋ	m, n, ŋ

(b) σ = 0.01; n. ev. = -0.25 (“Don't generalize”)		
Input:	m, ŋ	m, n, ŋ
Output:	m, ŋ	m, n, ŋ
	m, ŋ	m, n, ŋ
	m, ŋ	m, n, ŋ
	m, ŋ	m, n, ŋ
	m, ŋ	m, n, ŋ

(c) σ = 0.25; neg. ev. = -0.15 (“Generalize”)		
Input:	p, m	p, b, m
Output:	p, b, m	p, b, m
	p, b, m	p, b, m
	p, b, m	p, b, m
	p, b, m, n	p, b, m
	p, b, m	p, b, m

(d) σ = 0.8; neg. ev. = -0.15 (“Generalize”)	
Input:	s, ʃ, tʃ
Output:	s, ʃ, tʃ, z, ʒ, dʒ
	s, ʃ, tʃ, z, ʒ, dʒ
	s, ʃ, tʃ
	s, ʃ, tʃ, z, ʒ, dʒ
	s, ʃ, tʃ, z, ʒ, dʒ, f, ɔ

Table 2. (a) Emerging nasals; (b) stable nasals; (c) emerging labials; (d) emerging sibilants.

Table 2(c) shows a similar case in which a /b/ gap in the phonetically unnatural class /p m/ tends to be filled, creating the class of labials, while the class of labials is more stable. In Table 2(d), voiceless sibilants generalize to the class of all sibilants in three out of five trials. In one case they remain stable, and in one case the overgeneralization includes a marginal sibilant ([f]) and a vowel. These examples show that overgeneralizations, when they do occur, tend to favor natural classes. The isolated unnatural innovative class, such as [s ʃ tʃ z ʒ dʒ f ɔ], seems unlikely to become dominant in a community where they are outnumbered by natural innovative classes.

4. Conclusions

The simulation has shown that natural classes emerge in a model that has access to the observable phonetic properties of sounds, but no innate features. Innate features are not needed to rule out phonetically unnatural classes, because their relative rarity is already accounted for. Further, unnatural classes *do* occur in reality, so ruling them out is inherently problematic. The situation suggested by the simulation, in which any class is possible but phonetically natural classes are favored, is consistent with the survey results, which found a preference for phonetically natural classes but no categorical division of classes into natural and unnatural. The actual distribution of classes can potentially be accounted for in terms of the way sound patterns originate and change over time. Features can be posited *in response* to classes confronted by the language learner. Finally, even if innate features exist, this type of approach is still necessary in order to account for why some formally equivalent classes tend to be more common than others.

References

Blevins, Juliette. 2004. *Evolutionary Phonology*. Cambridge: Cambridge University Press.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. Cambridge, Mass.: MIT Press.

Clements, G.N. and Elizabeth V. Hume. 1995. The Internal Organization of Speech Sounds. In John Goldsmith, editor, The Handbook of Phonological Theory. Blackwell, Cambridge Mass., pages 245–306.

Dolbey, Andrew E. and Gunnar Ólafur Hansson. 1999. The source of naturalness in synchronic phonology. In Sabrina J. Billings and John P. Boyle and Aaron M. Griffith, editors, *CLS 35, Vol. 1*. CLS: Chicago, pages 59-69.

Frisch, Stefan, Janet Pierrehumbert and Michael Broe. 1997. Similarity and Phonotactics in Arabic. Indiana University ms., ROA-223.

Holmes, John and Wendy Holmes. 2001. *Speech Synthesis and Recognition*, 2nd Edition. New York: Taylor & Francis.

Hyman, Larry. 1977. Phonologization. In A. Juilland, editor, *Linguistic studies presented to Joseph H. Greenberg*. Anna Libri, Saratoga, pages 407–418.

Jakobson, Roman, C. Gunnar M. Fant and Morris Halle. 1954. *Preliminaries to speech analysis: the distinctive features and their correlates*. Cambridge, Mass.: MIT Press.

Janda, Richard D. 1999. Accounts of phonemic split have been exaggerated - but not enough. In *Proceedings of ICPhS 14*.

McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In Pat Keating, editor, Phonological Structure and Phonetic Form: *Papers in Laboratory Phonology III*. Cambridge University Press, Cambridge, pages 191–251.

Mielke, Jeff. 2004. The Emergence of Distinctive Features. Ph.D. thesis, The Ohio State University.

Mielke, Jeff, Adam Baker, Diana Archangeli and Sumayya Racy. 2005. Palatron: a technique for aligning ultrasound images of the tongue and palate. In Scott Jackson and Daniel Siddiqi, editors, *Coyote Papers vol. 14*.

Odden, David. 2002. The verbal tone system of Zina Kotoko. In Schmidt, Odden and Homberg, editors, *Aspects of Zina Kotoko grammar*. Lincom Europa, München.

Proceedings of the 24th West Coast Conference on Formal Linguistics

edited by John Alderete,
Chung-hye Han, and Alexei Kochetov

Cascadilla Proceedings Project Somerville, MA 2005

Copyright information

ISBN 1-57473-407-5 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Mielke, Jeff. 2005. Modeling Distinctive Feature Emergence. In *Proceedings of the 24th West Coast Conference on Formal Linguistics*, ed. John Alderete et al., 281-289. Somerville, MA: Cascadilla Proceedings Project.

or:

Mielke, Jeff. 2005. Modeling Distinctive Feature Emergence. In *Proceedings of the 24th West Coast Conference on Formal Linguistics*, ed. John Alderete et al., 281-289. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1233.