ABSTRACT

The basic representational hypothesis in phonology is that segments are coded using a universal set of discrete features. We propose a method for quantitatively measuring how well such features align with arbitrary segment representations. We assess articulatory, spectral, and phonotactic representations of English consonants. Our procedure constructs a concrete representation of a feature in terms of the pairs it distinguishes, and can be extended to any pair of representations to test the consistency of one with the individual dimensions of the other. We validate the method on our phonetic representations and then show that major natural classes are not well represented in the surface phonotactics.

Keywords: features; perception-production; quantitative methods; phonotactics

1. INTRODUCTION

Where do phonological features come from? Most feature sets attempt to make the classes that are “natural” to a learner featurally simpler; but it is also assumed that features are phonetically grounded. To assess how well natural classes are grounded in the phonetics, previous work has taken representations of segments based on articulation, acoustics, and tendency to group together crosslinguistically, and compared them qualitatively using visual inspection of a principal component analysis [7].

We propose a method for assessing this kind of alignment quantitatively. We limit ourselves to alignment with hand-defined natural classes, and demonstrate its use in finding what is coded in rich acoustic and articulatory representations. We then apply the method to the question of whether phonological features are emergent, i.e., substance-free and grounded only in language-specific phonological patterning [9, 6]. This would imply that phonological patterning should determine, and therefore align with, the phonological features. We find no evidence in favor of this hypothesis as it applies to English surface phonotactics aligning with standard place, manner, and voicing features on consonants.

2. DATA

To develop and test our method, we compared acoustic, articulatory, and phonotactically based representations of the consonant inventory of English. The acoustic representation of a phone was the corpus-wise average of 40 mel scaled filter banks, in a stack of 11 frames centered on the midpoint of each token of the category (25 ms windows spaced at 10 ms intervals), taken from TIMIT [3]. The articulatory representation of a segment was taken from the data reported on in [7]: vertical oral cavity distances estimated from ultrasound, plus vocal fold activity from EGG and oral and nasal airflow measurements, averaged over the productions of three trained phoneticians in V–V context for each of the three corner vowels. See [7] for details.

We used two different methods to extract a phonotactic representation for a segment without using features, both applied to the phonetic transcripts from the naturalistic Buckeye interview corpus [10]. The first uses a neural network (NN) approach where the surrounding phone sequence (plus and minus two segments) have to be predicted from a central phoneme [8]. The NN is a log–linear model composed of an input layer (of 39 phonemes) mapping to an embedding (of dimension 10) and then to the output layer (logistic regressions to 4^39 phonemes), and is trained with backpropagation with the logistic loss. The weights are used as a representation for the input phone. The second uses a singular value decomposition approach (SVD). We recorded relative frequencies of individual left-context phones and bigrams, right-context phones and bigrams, and of individual left-and-right contexts with a window of both one and two segments. We took the resulting table of relative frequencies and applied matrix factorization to reduce the dimension to 30. Both models ignored word boundaries. (An SVD model that included word boundaries was not meaningfully
3. METHOD

3.1. Constructing representations of features

Table 1:
3.2. Comparing representations of features

\[ \text{alignment} \]

\[
\begin{pmatrix}
 f & f & f \\
 f & f & f
\end{pmatrix}
\]

\[
\begin{pmatrix}
 0 & 0.5 \\
 0.5 & 1
\end{pmatrix}
\]

\[
\begin{pmatrix}
 f & f & f \\
 f & f & g
\end{pmatrix}
\]

4. RESULTS
Figure 4:

Alignment scores for all four test representations (0.5 is chance and 1.0 is the maximum) versus manner features ([continuant] and [nasal]), place contrasts ([coronal]–[dorsal], [coronal]–[labial], and [labial]–[dorsal]), and the feature [voice].

The test space similarity number of frep pairs for good (upper and lower left) versus badly performing test representations (upper and lower right).

5. CONCLUSION

We presented a new method for evaluating the grounding of phonological representation in acoustic, articulatory and phonotactic space. The results on acoustic and articulatory are consistent with expectations about the grounding of place and manner features. The fact that the phonotactics of English align so poorly with basic natural classes prima facie challenges emergent feature theories. We have certainly not exhausted the space of cognitively reasonable phonological pattern based representations; nevertheless, given that at least some of the features we tested are phonologically active in English (for example, place, in nasal place assimilation, and voicing, in coda obstruent voicing restrictions), it is worrying that we do not see them robustly attested.

It is unsurprising that at least the core of the standard phonological feature system for consonants is consistent with major dimensions of phonetic variability, since standard phonological features are at least partly phonetically interpretable by design. The general procedure we have presented does not allow us to evaluate the hypothesis that any of these features is cognitively active; rather, it allows us to assess which sources of information are available in acquisition and processing, if any, that would support their being cognitively active. That we have a quantitative measure of this is an advance over previous research.

Future work should explore general procedures for working with a classifying representation to automatically determine the dimensions of greatest contrast, keeping in mind that this may be different depending on a segment’s location in the classifying space. With such a tool available, it would become straightforward to compare arbitrary pairs of representations. An extension to continuous classifying representations can be made, for example, to sum-
6. REFERENCES

The contrastive hierarchy in phonology

TIMIT: acoustic-phonetic continuous speech corpus

Linguistic inquiry

Psychological review
The emergence of distinctive features

Lingua

Advances in Neural Information Processing Systems

Lingua

Speech Communication