Surveying Auditory Space Using Vowel Formant Data

Matthew J. Makashay* and Keith Johnson†

Department of Linguistics, Ohio State University, 222 Oxley Hall, Columbus, Ohio 43210

Abstract: Auditory neural map (ANM) formation through unsupervised learning of Gaussian distributions around average vowel formants appears to produce clusters in the ANM which demonstrate the perceptual magnet effect (2) and could serve as a basis for vowel categorization (1). The simulations reported in this paper tested this hypothesis by examining the maps produced by training on average male vowel formants, raw male formants, and average male and female formants and F0. Results showed that ANM clusters do not correspond to linguistic vowel categories for multiple talkers.

Unsupervised learning of vowels in an ANM warps the perceptual vowel space in a way consistent with the perceptual magnet effect (2,1). The resulting auditory maps contain vowel clusters which appear to provide a good basis for the perceptual categorization of vowels. However, the previous simulations (1) were based on Gaussian distributions centered on average vowel formants. The simulations described in this paper replicate and extend the earlier work in (1) by examining clustering in ANMs produced through training on a wider range of variation. The purpose was to test the hypothesis that unsupervised ANM formation provides a basis for vowel categorization.

METHODS

Random values based on a Gaussian distribution of F1 and F2 Mel values taken from eight steady-state English vowels were presented to an ANM following the methods described in (1). This training was unsupervised, as the inputs were not labeled by vowel. After training had completed, the preferred stimuli (i.e. the input to which each map location is most highly attuned) were plotted to find any evidence of clustering. Statistical analysis using K-means cluster analysis was also employed to find evidence of data aggregation.

MODELING

Average male F1 and F2 values for eight vowels (4) were used in the first simulation for the single talker case. As shown in figure 1, results are very similar to those found in (1) for the modeling of the perceptual magnet effect using a Gaussian distribution of average F1 and F2 values. These perceived values display clustering about the vowel category centers, and thus may serve as an auditory basis for linguistic vowel categorization.

Average male and female F0, F1, and F2 values for eight vowels (4) were tested in the second simulation. Figure 2 shows that male and female vowels remained separate in the ANM. The distinctions seen in the F1/F2 plane were further distanced across two planes in the F0 dimension (not shown), with one F0 plane for the male talkers, and another for the female talkers. However, figure 3 shows these formants will produce clear clusters when the input values are normalized (3).

The third simulation trained an ANM using F1 and F2 values from two tokens of each of eight steady-state vowels from 14 male native speakers from central Ohio, for a total of 244 tokens. The resulting preferred formant values for each map location is shown in figure 4. In this case, the vowel categories are less distinct, larger, and less regularly shaped than in the average vowel condition, even when normalized (figure 5).

DISCUSSION

In unsupervised auditory map formation, natural variation among talkers in simulation 3 resulted in a loss of distinction among the vowel categories in the ANM. This was caused by the realization of different tokens for

2037
speakers' vowel categories, such as one talker's [æ] formants encroaching the vowel space of another talker's [ɛ] formants. Natural variation among male talkers in simulation 3 led to a merging of clusters, while average male and female talker variation in simulation 2 led to an over-abundance of clusters if formants are not normalized. These findings suggest that although auditory neural map formation may provide an elegant account of the perceptual magnet effect, clusters in the ANM do not correspond to linguistic vowel categories when the full range of acoustic variation found in speech is considered.

ACKNOWLEDGEMENTS

Work supported by *NSF Graduate Fellowship and †NIH FIRST Award R29-DC01645-06.

REFERENCES