(2004), Rose & Walker (2004), Coetzee & Pater (2005), Herd (2005), Mackenzie (2005), Bailey & Hahn (2005), Kessler (2005), and many others.[1] Data on the phenomenon of similarity avoidance was gathered in an experiment on East Bengali echo reduplication discussed below. But how is similarity actually calculated? To answer this question, four theories of similarity were tested against the experimental data and compared to one another.

1.2 The alternation: East Bengali fixed-segment echo reduplication

Fixed-segment reduplication involves copying all base material into the reduplicant, except for one part, which is replaced with a fixed segment (FS) (McCarthy & Prince 1986, Nevins & Wagner 2001); echo reduplication is one instantiation of this process. The default East Bengali echo reduplication pattern is shown in (1) and (2), where the reduplicant-initial segment is usually replaced with default fixed segment /t/:[2]

(1)	pani	'water'	(2)	kaʃi	'cough(s)'
	pani tani[3]	'water, etc.'		kaʃ	

2 Experiment

2.1 Research question

Having briefly described the phenomenon of similarity avoidance, it is nevertheless unclear on what basis speakers are judging similarity. Do features and natural classes play a role? Do patterns in the lexicon play a role? Does the phoneme inventory play a role? To better understand what factors determine consonant similarity, an experiment was carried out with the purpose of gathering data on echo reduplication using productions of native speakers. Using this data, four theories of similarity were tested against the observed patterns.

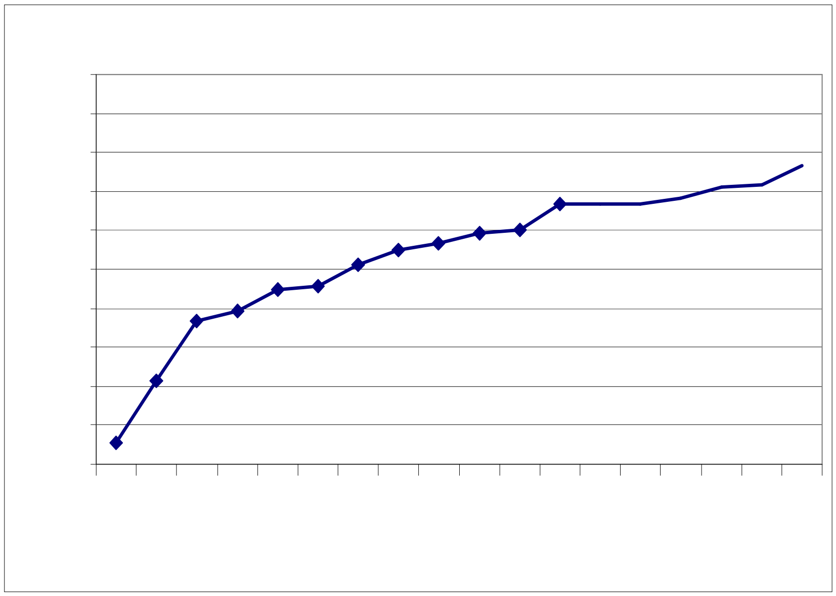
2.2 Methods

Thirty (30) adult native speakers of Bengali were presented auditorily with recordings of 60 native Bengali disyllabic roots, grouped by their initial consonant.[4] These included eight (8) stimuli beginning with /t/ (i.e. the identity condition), 23 stimuli beginning with consonants potentially considered similar to /t/ – /tʰ, d, t̪, t̪ʰ, tʃ/ – (i.e. the similarity condition), and 29 stimuli beginning with other consonants (i.e. the control condition). No word included consonants from the similarity condition (i.e. /tʰ, d, t̪, t̪ʰ, tʃ/) in non-initial position.

The stimuli were produced in two dialects spoken in urban Bangladesh (i.e. Standard Bengali and East Bengali)[5] by an adult female speaker in a sound-proof booth. The order of stimuli was randomized for each subject. After the stimulus was played aloud to subject (who chose the dialect in which to hear the stimuli), the subject was asked to repeat the word aloud with its reduplicant.

2.3 Results

The experimental results confirm that the overall pattern of echo reduplication exhibits both identity- and similarity avoidance. Bases with initial consonants such as /t, tʰ, d, t̪/ took very few reduplicants with fixed segment /t/, while bases with initial consonants such as /l, m, p, bʱ/ most often took reduplicants with fixed segment /t/. Bases with other initial consonants – those of intermediate similarity to /t/ – showed more variable behavior. As shown in Figure 1, the percentage of fixed segment /t/-use in echo reduplicants is inversely related to the presumed similarity between /t/ and the base-initial consonant.



3.1.1 Implementation

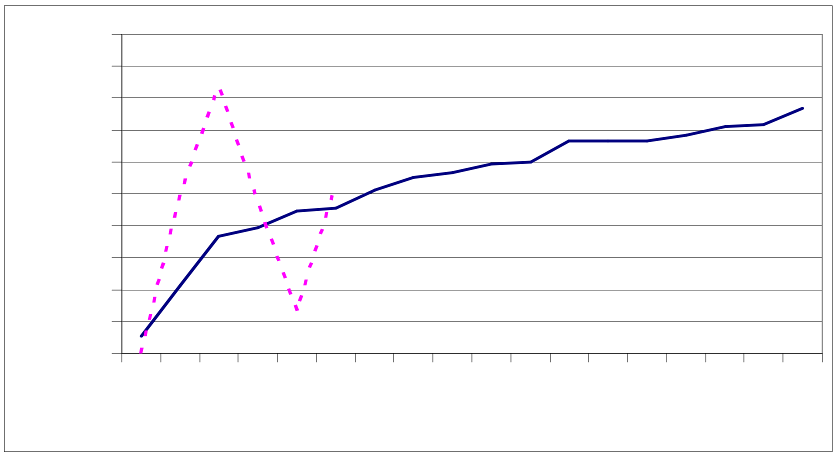
To test this theory, the cooccurrence of /t/ with each consonant (C) in roots of the shape /tVCV/ and /CVtV/ was calculated as in (9), using phoneme distribution data from Mallik et al. (1998). The numerator represents observed cooccurrence and the denominator represents expected cooccurrence.

(9)

$$\frac{\dfrac{\text{Observed } \{ C, t \} \text{ cooccurrence in roots}}{\text{Total roots}}}{\dfrac{\text{Observed /C/ occurrence in roots}}{\text{Total roots}} \times \dfrac{\text{Observed /t/ occurrence in roots}}{\text{Total roots}}}$$

If /t/ and a consonant C cooccur with an Observed/Expected (O/E) value less than 1, it is likely that the two consonants are subject to a cooccurrence

examples of incorrect predictions are circled in Figure 3. Note how Theory II predicts that /t̪/ is most similar to /t/, followed by /tʃ/, /d/, and then /tʰ/, while the data suggests that /tʰ/ is most similar to /t/, followed by /d/, /t̪/, and then /tʃ/.



similarity of two consonants, certain features are more heavily weighted than

make no predictions about similarity phenomena in other languages. It is unclear

Given the relatively crowded articulatory space occupied by these 15 obstruents, it is reasonable to postulate that speakers of such a language stretch the perceptual space between those phonemes (Kuhl 1991, Kuhl 2000, Iverson et al. 2003) by amplifying the importance of each relevant feature. If this data is representative of a larger pattern, we can predict that while phonetic features are universally available, they have language-specific weights derived from the phoneme inventory, with each weight corresponding to the capacity of each feature to make phonemic contrasts. For example, since the feature [voice] alone distinguishes ten pairs of consonants – more than any other feature in the language – it is not surprising that it is assigned the heaviest weight in the language (w = 0.554). Under this hypothesis, speakers acquire the feature weights of their language once they acquire the full phonemic inventory, and are then equipped to make similarity judgments in the productive grammar.

6 Conclusion

Numerous measurements of similarity have been proposed for a variety of languages and processes, making reference to the lexicon, universal features, and OT constraints. Four theories of similarity were tested against data collected in an experiment studying a productive similarity avoidance alternation (i.e. East Bengali echo reduplication). The theory that best matched the observed data involved feature weighting: speakers measure the similarity of consonants by referring to the features they share, counting certain features as having heavier weights. One hypothesis on the source of these weights involves the concept of contrast: a feature's weight is determined by its ability to contrast phonemes in the inventory. This suggests that similarity is measured using universally-available features assigned weights reflecting their relative effectiveness in contrasting the phonemes in the inventory. Data from productive similarity avoidance alternations in several languages will be needed to test this further.

Notes

1 I would especially like to thank my M.A. thesis advisors, Kie Ross Zuraw, Colin Wilson (also my programming and statistics consultant), and Bruce Hayes; my native speaker consultant, Farida Amin Khan; the UCLA Phonology Seminar; and the 30 subjects of my study.

2 All examples shown in the current study were collected in the experiment described in Section 2.

3 All fixed segments are shown in boldface to distinguish them from surrounding material.

4 See Khan (2006) for a full list of all stimuli used in the experiment described in Section 2.

5 Bengali is an Indo-European language spoken by over 171 million people in South Asia (Gordon 2005). Based on the dialect of towns near Dhaka, East Bengali is widely understood by speakers of other Bangladeshi dialects, although Standard Bengali is the only form used in schools or the media.

Frisch, Stefan. 1996. 'Similarity and frequency in phonology', Unpublished Ph.D. thesis, Northwestern University.

Frisch, Stefan, Janet Pierrehumbert, & Michael Broe. 2004. 'Similarity Avoidance and the OCP', McCarthy, John, & Alan Prince. 1986. 'Prosodic Morphology', Ms., University of Massachusetts – Amherst & Brandeis University.

Melnar, Lynette & Chen Liu. 2006. 'A Combined Phonetic-Phonological Approach to Estimating Cross-Language Phoneme Similarity in an ASR Environment', Proceedings of the Eighth Meeting of the ACL Special Interest Group on Computational Phonology at HLT-NAACL 2006. New York: Association for Computational Linguistics. 1-10.

Nevins, Andrew & Michael Wagner. 2001. 'Backcopying, Shmackcopying: Correspondence Relations and Dissimilation in Echo Reduplication', Presented at the LOT Summer School, Nijmegen. Ms.

Pierrehumbert, Janet B. 1993. 'Dissimilarity in the Arabic Verbal Roots', Proceedings of the North East Linguistics SocDml [,eetSzq,iSzq, 0Sq,ASF0 c,CSq,Dmlzq, δmSzq,.Szq,y,eSzD0,sSqDdkk0S&G,t, n Sz6,iSzSI