Similarity Avoidance in East Bengali Fixed-Segment Echo Reduplication

Sameer ud Dowla Khan
UC Los Angeles

Abstract

Many languages employ reduplication processes in which one segment of the reduplicant is fixed. In East Bengali, /**t**/ is normally the initial consonant in reduplicants of this type (/gaɹi-**t**aɹi/ 'cars, *etc.*'). However, when the base itself begins with /t/, speakers prefer alternate fixed segments, such as /**f**/ or /**m**/ (not */tægɹa-**t**ægɹa/, but /tægɹa-**m**ægɹa/ 'cross-eyed, *etc.*'). Speakers also tend to avoid fixed segment /**t**/ when the base begins with a consonant that is similar to /t/ (not ??/tʰoŋga **t**oŋga/, but /tʰoŋga **f**oŋga/ 'bags, *etc.*'). Results of an experiment indicate that bases beginning with consonants similar to /t/ (*e.g.* /tʰ/ or /d/) take reduplicants with fixed segment /**t**/ far less often than bases beginning with consonants dissimilar to /t/ (*e.g.* /p/ or /bʱ/). If the avoidance of similarity between the initial consonants of base and reduplicant is indeed at work, on what basis is similarity being measured? To determine consonant similarity, speakers could be accessing language-specific measures of consonant similarity by observing the patterns in their lexicon, or they could be using a more universal measure of similarity, based on the phonological features of each consonant. To better understand the measurement of similarity, four metrics were tested against the experimental data. The current study explores a simple model in which similarity is determined by the weighted sum of shared features. One hypothesis regarding the source of these weights is considered here – a feature's weight may reflect its ability to contrast the phonemes of the language.

1 Introduction

1.1 The question: how is similarity calculated?

The concept of similarity has been shown to be phonologically important in numerous studies, including Frisch (1996), Frisch, Pierrehumbert, & Broe

(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 **t**ani[3]	'water, *etc.*'		kaʃi **t**aʃi	'coughs, *etc.*'

However, alternate fixed segments (*e.g.* /**f**/, /**m**/, /**z**/, /**p**/, /**b**/) are also attested, as in (3) and (4) below:

(3)	tika	'vaccine(s)'	(4)	tʃakɹi	'career(s)'
	tika **f**ika	'vaccines, *etc.*'		tʃakɹi **b**akɹi	'careers, *etc.*'

1.3 The phenomena: identity- and similarity avoidance

The choice of fixed segment is subject to two restrictions: identity avoidance and similarity avoidance. Identity avoidance is the rejection of reduplicants with a fixed segment identical to the segment being replaced, as in (5) and (6):

(5)	tipːa	'having pressed'	(6)	muɹi	'puffed rice'
	*tipːa **t**ipːa	'h. pressed, *etc.*'		muɹi **t**uɹi	'puffed rice, *etc.*'
	tipːa **m**ipːa	'h. pressed, *etc.*'		*muɹi **m**uɹi	'puffed rice, *etc.*'

Speakers also tend to reject reduplicants with a fixed segment merely similar to the segment being replaced (*i.e.* similarity avoidance), as in (7) and (8):

(7)	t^{h}æka	'obstacle(s)'	(8)	t̪ala	'lock(s)'
	*t^{h}æka **t**æka	'obstacles, *etc.*'		??t̪ala **t**ala	'locks, *etc.*'
	t^{h}æka **f**æka	'obstacles, *etc.*'		t̪ala **p**ala	'locks, *etc.*'
	t^{h}æka **m**æka	'obstacles, *etc.*'		t̪ala **m**ala	'locks, *etc.*'

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.

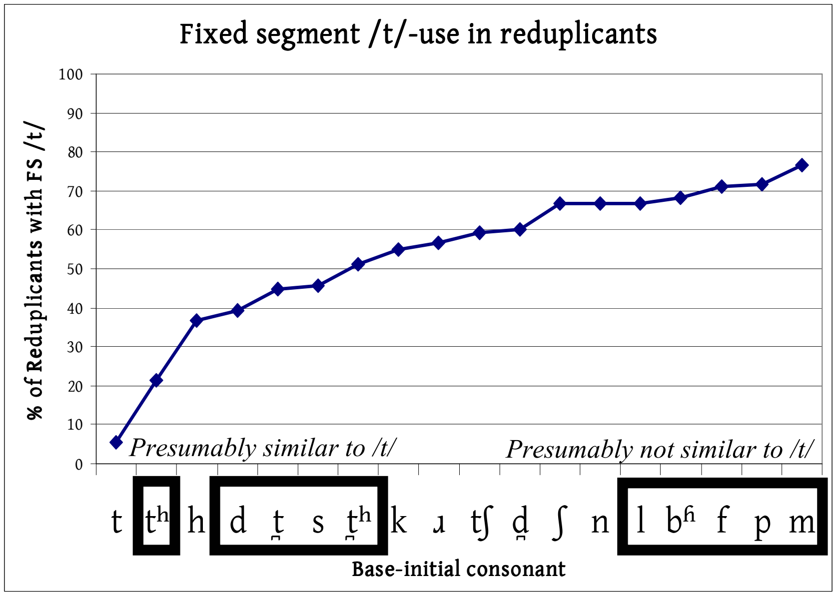

Figure 1. Fixed segment /**t**/-use in reduplicants, arranged by base-initial consonant.

3 Analysis

What theory could explain the experimental data? Four theories of similarity are considered: lexical cooccurrence restrictions (OCP), the shared natural classes metric, relativized OCP constraints, and feature weighting.

3.1 Theory I: lexical cooccurrence restrictions (OCP)

The Obligatory Contour Principle (OCP) describes the tendency of identical (and similar) consonants to cooccur less frequently than more dissimilar consonants within roots in a given lexicon (McCarthy 1986, Pierrehumbert 1993). If these OCP restrictions in the lexicon are the only constraints penalizing similarity in the productive grammar, then speakers could infer similarity values from cooccurrence rates in the lexicon – or, similarity values might come from some other source, but still be reflected in both lexical cooccurrence rates and reduplicative behavior. Thus, bases with initial consonants that cooccur less often with /t/ in the lexicon should allow fewer /**t**/-reduplicants than other bases. If this prediction is borne out, we can conclude that similarity is a potentially language-specific measure, based on or at least related to patterns in the lexicon.

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

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 restriction, and are thus considered more similar to each other in the language. An O/E value greater than 1 suggests that /t/ and the consonant C are not subject to a cooccurrence restriction, and are being treated as less similar to each other.

3.1.2 Comparison with results

Figure 2 compares O/E values (multiplied by a constant 30) to the observed data.

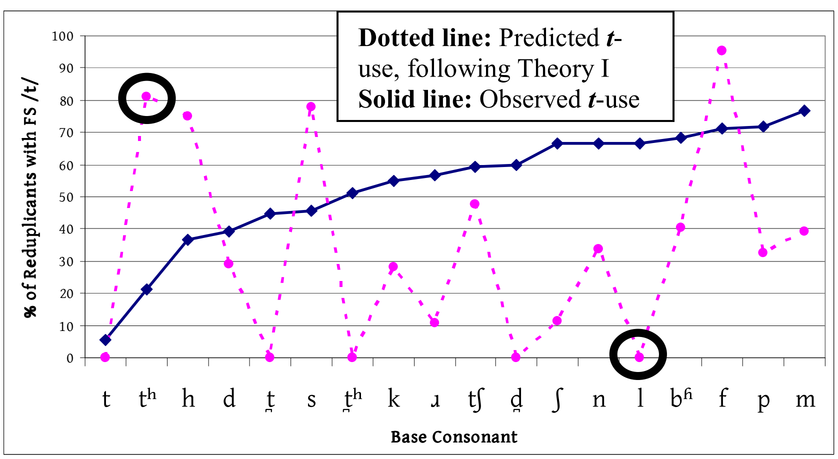

Figure 2. Fixed segment /**t**/-use in reduplicants as predicted by Theory I/lexical cooccurrence restrictions (dotted) versus observed data (solid).

As illustrated in Figure 2, there is no correlation between the predictions of Theory I (lexical cooccurrence restrictions) and the experimental data [r^2 = .004, p = 0.81]. Two points in Figure 2 are circled as illustrations of inaccurate predictions made by Theory I. Bases with initial /tʰ/ are wrongly predicted to almost exclusively take /**t**/ as the fixed segment, due to the large number of words like /tʰõte/ 'lips (LOC)', with /tʰ/ and /t/ cooccurring within the same root, while experimental data such as (10) show that /tʰ/-initial bases *do not* take fixed segment /**t**/ over 78% of the time. Conversely, bases with initial /l/ are wrongly predicted to never take /**t**/ as the fixed segment in echo reduplicants, due to the lack of roots of the shape /lVtV/. In the experimental data, over 65% of /l/-initial bases in fact do take fixed segment /**t**/, as in (11).

(10)	tʰoka	'knock(s)'	(11)	loha	'iron (Fe)'
	*tʰoka **t**oka	'knocks, *etc*.'		loha **t**oha	'iron, *etc*.'

The lack of correlation between Theory I and the observed data strongly suggests that the cooccurrence restrictions present in the Bengali lexicon are unrelated to the cooccurrence restrictions seen in echo reduplication.

3.2 Theory II: shared natural classes metric

Frisch (1996) proposes a similarity metric that counts the number of natural classes shared by two sounds, as in (12).[6] This metric was shown in Frisch *et al.* (2004) and subsequent studies to describe the lexicon and grammar of Arabic.

3.2.1 Implementation

The similarity score of each consonant with /t/ was calculated in the software program Similar.exe (Zuraw, n.d.) using the following equation:

(12) $$\text{Similarity } \{/t/, C\} = \frac{\text{\# of natural classes containing both /t/ and C}^{7}}{\text{\# of natural classes containing /t/ and/or C}}$$

Higher similarity scores (*i.e.* approaching 1) indicate more similar consonants, while lower scores (*i.e.* approaching 0) indicate less similar consonants. Following this metric of similarity, bases with initial consonants that share more natural classes with /t/ should take fewer /t/-reduplicants than other bases. This would suggest that similarity measurement has both a universal component (*i.e.* features) and a language-specific component (*i.e.* the phoneme inventory).

3.2.2 Comparison with results

The shared natural classes metric is better than Theory I at predicting most of the experimental results [r^2 = .584, $p < 0.01$]. However, it cannot predict the contrasts among the consonants in the similarity condition; the most striking 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/.

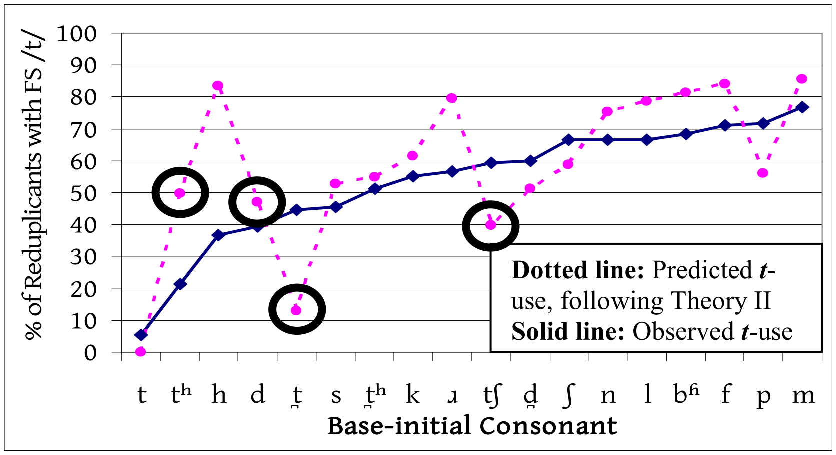

Figure 3. Fixed segment /**t**/-use in reduplicants as predicted by Theory II/shared natural classes metric (dotted) versus observed data (solid).

3.3 Theory III: relativized OCP constraints

Coetzee & Pater (2005) proposes an Optimality Theoretic (Prince & Smolensky 1993) account of Muna, where OCP constraints against certain larger feature combinations (*e.g.* OCP-LAB [αCNT][αSON]) are ranked above OCP constraints against smaller combinations thereof (*e.g.* OCP-LAB [αSON]); these are then ranked above a general OCP constraint (*e.g.* OCP-LAB). This ranking derives from the lexicon by an algorithm based on the type frequency of lexical exceptions to weak OCP constraints, and is used to describe lexical OCP restrictions. Applying this measure of similarity to Bengali, bases with initial consonants that share more combinations of features with /t/ should take fewer /**t**/-reduplicants than other bases (although, of course, the ranking would not derive from the Bengali lexicon, as it was shown in Section 3.1.2 that lexical patterns are a poor model of reduplicative phenomena). Like Theory II/shared natural classes metric, this theory would suggest that the measurement of similarity has both a universal component (*i.e.* the features) and a potentially language-specific component (*i.e.* the lexicon).

3.3.1 Implementation

As formulated in Coetzee & Pater (2005), the relativized OCP constraint ranking derives from the lexicon. However, since cooccurrence rates in the lexicon were found to be not correlated with the reduplication facts (see Section 3.1.2), such a ranking would be useless. Thus, the relativized OCP constraint hierarchy in (11) is directly fitted to the reduplication data making no reference to the lexicon:[8]

(13)	OCP-COR (αs.g., αvoi., αdist., αdel.rel., αant., αson., αnas., αlat.) »
OCP-COR (αvoi., αdist., αdel.rel., αant., αson., αnas., αlat.) »
OCP-COR (αdist., αdel.rel., αant., αson., αnas., αlat.) »
OCP-COR (αdel.rel., αant., αson., αnas., αlat.) »
OCP-COR (αant., αson., αnas., αlat.) »
OCP-COR (αson., αnas., αlat.) »
OCP-COR (αnas., αlat.) »
OCP-COR (αlat.) »
OCP-COR

3.3.2 Comparison with results

Theory III, as modified to make no reference to the Bengali lexicon, is a relatively close match to the observed data [$r^2 = .717$, $p < 0.01$].

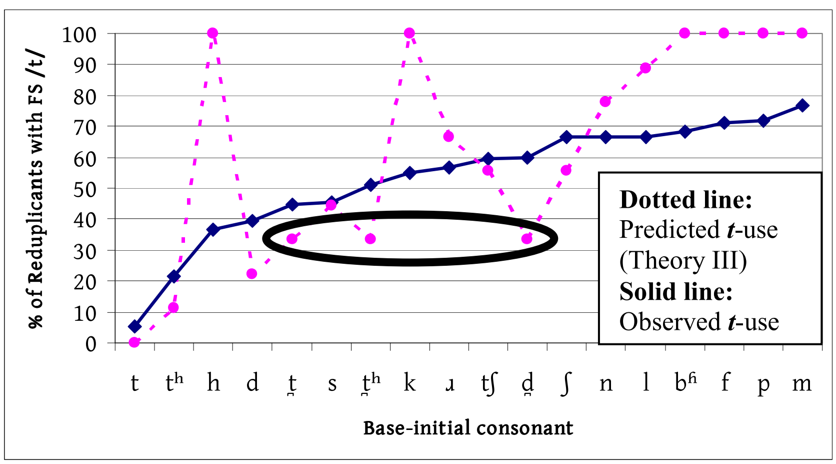

Figure 4. Fixed segment /**t**/-use in reduplicants as predicted by Theory III/relativized OCP constraints (dotted) versus observed data (solid).

Note, however, that all dentals /t̪, t̪ʰ, d̪/ are predicted by Theory III to be equally similar to /t/, when the data indicates /t/ is more similar to /t̪/ than to /t̪ʰ/ or /d̪/.

3.4 Theory IV: feature weighting

Consonant similarity may be measured by counting shared features. However, some features may be more important than others. Thus, when calculating the similarity of two consonants, certain features are more heavily weighted than others (Ladefoged 1969). By assuming that features can have unequal weights, we can predict that bases with initial consonants that share a greater total shared feature *weight* (as opposed to the raw *number* of features shared) with /t/ will take fewer /**t**/-reduplicants than other bases. If this prediction is borne out, we could conclude that the measurement of similarity has a universal component (*i.e.* features) while allowing for (possibly language-specific) weights.

3.4.1 Implementation

As the lexicon has been shown to be a poor source of similarity measurements in the results of Theory I, we can only gather information on consonant similarity from the reduplication data itself. Thus, the calculation of feature weights was performed using the software program R (R Development Core Team 2005) by fitting the *sim*(C, t) values in (14) to the observed values of P:[9]

(14) $$P = ((m!) / (n!(m - n)!))\,(1 - sim(C, t))^{n}\,(sim(C, t))^{m-n}$$

Where *P* is the probability that base-initial *C* will cooccur with default fixed segment /t/ *n* times out of *m* trials, and *sim*(C, t) was calculated as in (15):

(15) $$sim(C, t) = \exp(-\sum_{i=1}^{\#ftrs} w_i(1 - \delta_i(C, t)))$$

The weights found to be most effective in modeling the data are as follows: all but four features received the default weight (*w*) of 0.100. The four features that were found to have heavier weights were [voice] (w = 0.554), [distributed] (w = 0.400), [strident] (w = 0.249), and [spread glottis] (w = 0.198). These four features turn out to be independently important in the language. This is discussed further in Section 5.

3.4.2 Comparison with results

Feature weighting achieves the best match with the experimental data [$r^2 = .855$, $p < 0.01$], as shown in Figure 5.

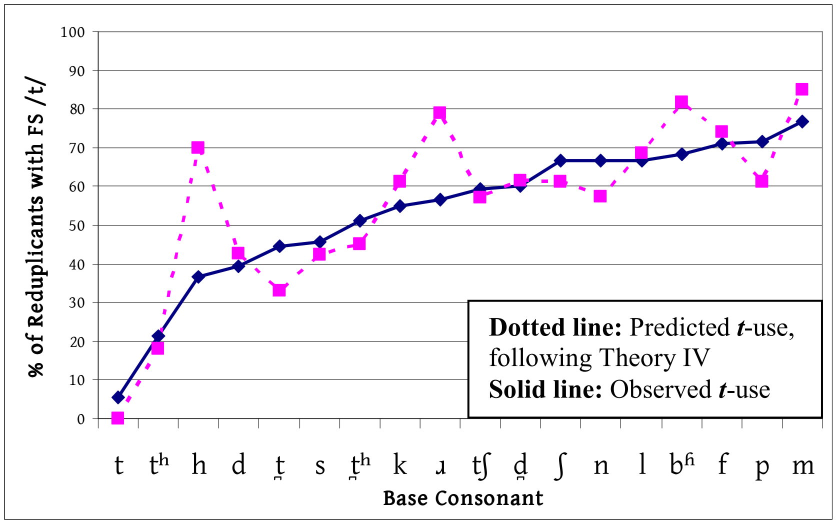

Figure 5. Fixed segment /t/-use in reduplicants as predicted by Theory IV/feature weighting (dotted) versus observed data (solid).

4 Summary

Theory	Derives from	Free parameters	Correlation with data
I. Lexical cooccur. restrict.	Lexicon	none	$r^2 = .004$ $p = 0.81$
II. Shared natural classes metric	Phoneme inventory	none	$r^2 = .584$ $p < 0.01$
III. Relativized OCP, as adapted	?	feature subsets	$r^2 = .717$ $p < 0.01$
IV. Feature weighting	? *(see below)*	ftr. weights *(see below)*	$r^2 = .855$ $p < 0.01$

Table 1. Summary of the comparison of four theories of similarity.

The lack of correlation between lexical cooccurrence restrictions and /t/-use in reduplicants [$r^2 = .004$] confirms that speakers do not judge similarity based on cooccurrence patterns in their lexicon. The correlation between shared natural classes and /t/-use in reduplicants [$r^2 = .584$] is substantial, but the model cannot describe the relative similarity of the coronal consonants to /t/. And while the data can be closely predicted by positing relativized OCP constraints [$r^2 = .717$], this requires the use of eight constraints that have no basis in the lexicon, and make no predictions about similarity phenomena in other languages. It is unclear how speakers could acquire a grammar involving these relativized OCP constraints from independent sources. The theory that best fits the data [$r^2 = .855$] is one in which similarity is judged based on universal features assigned different weights. Of course, in addition to being the closest matches to the observed patterns, the predictions of both Theory III/relativized OCP constraints and Theory IV/feature weighting were made to fit to the experimental data. Section 5 provides a discussion of one possible hypothesis supporting the theory that most closely matches the data (*i.e.* Theory IV/feature weighting).

5 Feature weights: a function of contrast?

The results of Theory IV bring up further questions: if similarity is judged by assigning different weights to different features, where could these weights come from? Weights could be either universal (Melnar & Liu 2006) or language-specific. If they are universal, we would predict that all languages would pattern with East Bengali and weight [voice], [distributed], [strident], and [spread glottis] above other features. This is unlikely, as many languages do not make contrasts using these features, unlike East Bengali.

Feature weights are more likely language-specific, possibly reflecting the relative importance of the feature in contrasting the language's phonemes. This hypothesis would predict that the better a feature is at contrasting the phonemes in a given language, the heavier its weight will be in that language. Thus, feature weights can serve as a function of contrast.

Consider the four features weighted more heavily than others in East Bengali echo reduplication: [voice], [distributed], [strident], and [spread glottis]. These four features alone can distinguish all 15 coronal obstruents from one another in East Bengali, and thus presumably carry significant practical import in terms of contrast in a language with such a coronal-heavy inventory.

	Dental	Alveolar	Post-alveolar
Plosive	t̪ d̪ t̪ʰ d̪ʱ	t d tʰ dʱ	
Affricate			tʃ dʒ tʃʰ dʒʱ
Fricative		s z	ʃ

Table 2. The coronal obstruents of East Bengali.

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.

6 Introduced in Pierrehumbert (1993); equation copied from Frisch, *et al.* (2004), example (7).

7 The feature values used in the current study are drawn from Hayes (2001), and represent the acoustic, articulatory (palatographic), and phonological data on a variety of Bengali spoken in Dhaka, Bangladesh, as presented in Hai (1960). See Khan (2006) for a table of feature values used.

8 This is the closest match between the relativized OCP theory and observed /**t**/-use. The ranking was derived as follows: since /tʰ/ is the second-most similar consonant to /t/ (after /t/ itself) in terms of /**t**/-use, the second highest-ranked constraint differs from the highest-ranked constraint in the loss of the feature [s.g.] (*i.e.* the feature contrasting /t/ and /tʰ/). Since the next most similar consonant to /t/ is /d/, the feature removed from the next constraint in the hierarchy is [voice], and so on.

9 The similarity of a consonant C and /t/ is calculated in the feature weighting metric using the following binomial formula:

$$P = ((m!) \div (n!(m-n)!))\ (1 - \mathrm{sim}(C, /t/))^{n}\ (\mathrm{sim}(C, /t/))^{m-n}$$

Where P is the probability that base-initial C will cooccur with FS /t/ n times out of m trials

Where *sim*(C, t) was calculated as:

$$sim(C, t) = \exp(-\sum_{i=1}^{\#\text{ features}} w_i(1 - \delta_i(C, /t/)))$$

Where $\delta_i(C, /t/) = 1$ if C and /t/ agree on feature i
$\delta_i(C, /t/) = 0$ otherwise

∴ $-\delta_i(C, /t/) = 0$ if C and /t/ agree on feature i
$-\delta_i(C, /t/) = 1$ otherwise

∴ $w_i\ \delta_i(C, /t/) = 0$ if C and /t/ agree on feature i
$w_i\ \delta_i(C, /t/) = w_i$ otherwise

∴ $\sum_{i=1}^{\#\text{ features}} w_i(1 - \delta_i(C, /t/))$ = sum of weights of features that distinguish C and /t/

∴ $0 \leq sim(C, /t/) \leq 1$

Where $sim(C, t) = 1 = \exp(-0)$ if C and /t/ are featurally identical
$sim(C, t) < 1$ if C and /t/ differ featurally

References

Bailey, Todd M. & Ulrike Hahn. 2005. 'Phoneme similarity and confusability', *Journal of Memory and Language,* **52**.

Berkley, Deborah. 1994. 'The OCP and gradient data', *Studies in the Linguistic Sciences,* **24**. 59-72.

Coetzee, Andries & Joe Pater. 2005. 'Gradient phonotactics in Muna and Optimality Theory', University of Michigan & University of Massachusetts – Amherst.

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', *Natural Language and Linguistic Theory,* **22**.

Gordon, Raymond G., Jr. (ed.). 2005. Ethnologue: Languages of the World, Fifteenth edition. Dallas, Tex.: SIL International. Online version: www.ethnologue.com.

Hai, Muhammad Abdul. 1960. *A Phonetic and Phonological Study of Nasals and Nasalization in Bengali*. Dacca, East Pakistan: Dacca University Press [now Dhaka, Bangladesh: Dhaka University Press].

Hayes, Bruce. 2001. *Introductory Phonology*. Unpublished textbook. UCLA.

Herd, Jonathon. 2005. 'Loanword adaptation and the evaluation of similarity', *Toronto Working Papers in Linguistics,* **24**.

Iverson, Paul, Patricia K. Kuhl, Reiko Akahane-Yamada, Eugen Diesch, Yoh'ich Tohkura, Andread Kettermann, & Claudia Siebert. 2003. 'A perceptual interference account of acquisition difficulties for non-native phonemes', *Cognition*, **87**. 1: B47–57.

Khan, Sameer ud Dowla. 2006. 'Similarity Avoidance in East Bengali Fixed-Segment Reduplication', Presented at the Western Conference on Linguistics, Fresno, CA, Ms. Online version: http://www.linguistics.ucla.edu/people/grads/sameer/publications.html.

Kuhl, Patricia K. 1991. 'Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not', *Perception & Psychophysics*, **50**. 93-107.

Kuhl Patricia K. 2000. 'Language, mind, and the brain: Experience alters perception', *The New Cognitive Neurosciences*, ed. M. S. Gazzaniga. Cambridge, MA: MIT Press. 99–115.

Ladefoged, Peter. 1969. 'The measurement of phonetic similarity', *Proceedings of the conference on Computational linguistics*, Sång Saby, Sweden.

Mackenzie, Sara. 2005. 'Similarity and Contrast in Consonant Harmony Systems', *Toronto Working Papers in Linguistics,* **24**.

Mallik, Bhakti Prasad, Nikhilesh Bhattacharya, Subhas Chandra Kundu, and Mina Dawn. 1998. *The Phonemic and Morphemic Frequencies of the Bengali Language*. Kolkata, India: The Asiatic Society.

McCarthy, John J. 1986. 'OCP Effects: Gemination and Antigemination', *Linguistic Inquiry* 17.

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 Society,* **23**. 367–381.

Prince, Alan & Paul Smolensky. 1993. 'Optimality theory: Constraint interaction in generative grammar', Ms. Rutgers University and University of Colorado, Boulder.

R Development Core Team. 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Rose, Sharon & Rachel Walker. 2004. 'A typology of consonant agreement as correspondence', *Language,* **80**.

Zuraw, Kie. Similar.exe. Software program.

Sameer ud Dowla Khan
UCLA Department of Linguistics
3125 Campbell Hall
Los Angeles, CA 90095
sameer@humnet.ucla.edu