

The phonetics and phonology of tone retraction in Shona*

Scott Myers

University of Texas at Austin

s.myers@mail.utexas.edu

July, 2000

In Shona, a Bantu language spoken mainly in Zimbabwe, if a word ending with a sequence of more than one high-toned syllables occurs before a word beginning with a high tone, the high tone of the first word is retracted off the final syllable of the first word: High-High ## High → High-Low ## High. Examples are given in (1) – (2). High tone is indicated by acute accent and low tone by the absence of any accent.

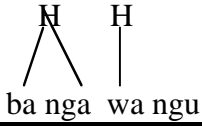
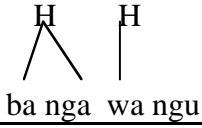
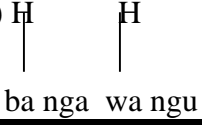
(1) (a) bá ngá	(b) bá ngá wángu
“knife”	knife my
	“my knife”

(2) (a) ku ón á	(b) ku ón a bángá
“to see”	to-see knife
	“to see a knife”

The retraction also occurs when comparable sequences occur within a word, as in (3).

* This research was made possible by a grant from the Office of the Vice-President for Research at the University of Texas at Austin. I would like to thank Augustine Agwuele for his help in digitizing the speech files, and Bob Ladd, Amalia Arvaniti, and Björn Lindblom for discussion of the experimental design. Thanks also to the Shona language consultants Theresa Kapungu, Timothy Nyangoni, Catherine Nyangoni, Stephanie Madziyauswa, Sunny Kapungu, and Jordan Maswoswe.

(5)

Input:	OCP	MAX-IO (A)
		
(a) 	*!	
(b) 		*

The first candidate is faithful to the input, but violates the OCP, since it includes two adjacent high tones. The second candidate violates Faithfulness, but not the OCP. Given the ranking, the one with retraction is optimal.

There is, however, an alternative analysis according to which the relevant process is not phonological but phonetic. A high tone is realized acoustically as a local peak in f_0 . It has been found in studies of a variety of languages that f_0 in a high-toned syllable begins to rise near the beginning of the syllable, but does not attain the local peak value until the end of that syllable or the beginning of the next (Silverman and Pierrehumbert 1990, Prieto, van Santen, and Hirschberg 1995, Arvaniti, Ladd and Mennen 1998, Myers 1999). The interval between the beginning of the high-toned syllable and the attainment of the f_0 peak is called *f_0 peak delay*. It has been found that, all else being equal, the f_0 peak occurs at a relatively constant proportion of the duration of the high-toned syllable.

The connection with tone retraction lies in the fact that f_0 peak delay has been found to be reduced when the following syllable bears a high tone. Silverman and Pierrehumbert (1990) found that the peak for a H* accent in English occurred significantly earlier if a following syllable was accented, and was earlier the closer that following accent was. Prieto et al. (1995) found the same for Spanish. But the effects are gradient, affecting where the f_0 peak is in the high-toned syllable, rather than affecting which syllable the high tone is associated with.

Could this be what is happening in Shona? Previous analyses of the pattern have all been based on transcription data, and such data is inherently categorical. Thus if the pattern is a gradient phonetic one rather than a categorical phonological one, we could never learn this from transcription data. There are many cases in which a pattern that has been treated as phonological on the basis of transcription data have turned out upon examination of instrumental data to be better treated as gradient phonetic patterns (Pierrehumbert and Beckman 1988, Cohn 1993, Myers 1999).

Thus the alternative analysis would be that there is no phonological retraction in these cases, so that the appropriate representation is the one to the left of the arrow in (4). What transcribers have heard as a tone retraction is the gradient retraction of the f_0 peak within the word-final syllable due to the high tone on the following syllable. The process is a matter of phonetic implementation rather than phonology. One point in favor of this view is Carter and Kahari's (1981) observation that their two speakers were "variable" in their use of the process, and tended not to apply it when the second word in the sequence began with a vowel.

To test this hypothesis, an experiment was performed in which f_0 timing was measured in the appropriate environment – at the end of a word before a word beginning with a high tone. In particular, the experiment addresses the question whether word-final HH is converted to HL when preceding a word-initial H. To answer this, the f_0 timing of underlying HH sequences in this context are compared with that in words ending underlyingly in HL. The hypothesis is that phonetic retraction renders the two tone classes more similar in this environment, but still distinct.

1. Materials and Procedures

Six speakers of Shona living in the Houston area participated in the study, all of whom are speakers of the Zezuru or Karanga dialects of Central Zimbabwe.

There were ten test sentences. All of the sentences contained a disyllabic high-toned verb stem. The consonants of the verb stems were always sonorants (liquids or nasals), since these disturb f_0 minimally, while providing clear boundaries for duration measurements. In one test sentence, the verb stem occurred in a recent past tense word, in which a disyllabic verb stem in the first person is always HL in tone. In a second test sentence, the same verb stem occurred in an infinitive form, in which a disyllabic high-toned stem always bears the tone pattern HH (Fortune 1985). These verbs were placed in carrier sentences with a preceding subject and a following context beginning with a high tone.

The test sentences were the following.

(6) (a) Iwé Kasadza wamára róri iyi.

you (name) you-past-scratch truck this

“Kasadza, you scratched this truck.”

(b) Handísí kumára róri iyi.

I-negative infinitive-scratch truck this

“I am not scratching this truck.”

(7) (a) Iwé Kasadza wanyíma Móry sadza.

you (name) you-past-refuse Mary porridge

“Kasadza, you refused Mary porridge”

(b) Handísí kunyíma Móry sadza

I-negative infinitive-refuse Mary porridge

“I am not refusing Mary porridge”

(8) (a) Iwé Kasadza warúma rúra iyi.

you (name) you-past-bite ruler this

“Kasadza, you bit this ruler”

(b) Handísí kurúmá rúra iyi.

I-negative infinitive-bite ruler this

“I am not biting this ruler”

(9) (a) Iwé Kasadza wanéna rúra iyi.

you (name) you-past-gnaw ruler this

“Kasadza, you gnawed this ruler”

(b) Handísí kunéna rúra iyi.

I-negative infinitive-gnaw ruler this

“I am not gnawing this ruler”

(10) (a) Iwé Kasadza wanyóra nyórwa murugwaro.

you (name) you-past-write writing in-page

“Kasadza, you wrote the writing in the page”

(b) Handísí kunyóra nyórwa murugwaro.

I-negative infinitive-write writing in-page

“I am not writing writing in the page”

The ten sentences were presented in a quasi-random order in a Powerpoint display on a laptop computer screen. Speakers read each sentence eight times. Half of the run-throughs of the list

were done at a self-selected normal rate, while in the other half speakers were instructed to speak “as fast as you can speak while still saying each word.” The sentences were recorded using a Sony DAT tape recorder and a high-quality Shure directional microphone.

The verb stems were either HL or HH in their isolation pattern. The difference in these two patterns lies in when f₀ falls: in HL we expect a fall from the first syllable to the second, while in HH the fall must be later. The timing of the f₀ fall was measured, using synchronized spectrogram and f₀ display in CSL. In particular, two measurements were made for each utterance: (a) *F0 Fall Delay* - the temporal distance between the onset of the stem-initial syllable (as indicated by a local minimum in amplitude) and the offset of the f₀ peak (i.e. the last f₀ point at the local maximum, after which f₀ begins to descend), and (b) *Syllable Duration* – the duration of the stem-initial syllable.

Some data had to be eliminated from the analysis due to problems of measurement, such as a gap at the crucial point in the f₀ track. For subject #5, all instances of the sentences in (6) and (7) had to be eliminated, since she consistently treated the words *-mara* and *Mary* as low-toned. This subject was also unfamiliar with the verb *-neŋa*, so sentence (9) had to be eliminated as well. This left for this subject only the data from the sentences in (8) and (10). All data from subject #6 had to be excluded, since he had all the problems of subject #5, and in addition spoke with frequent pauses and hesitations. For the other four subjects, all sentences were included.

The hypothesis is that HL and HH are not neutralized before H. This hypothesis would be supported if the timing of the f₀ fall with respect to the stem-initial syllable is significantly earlier in the case of HL than in the case of HH. If the two cases are not significantly different in this measure, this would support the phonological account in (4), according to which the retraction involves a categorical shift of the high tone off of the word-final syllable.

2. Results

The timing of the f0 fall with respect to the stem-initial syllable was regular, but there was no significant difference in this timing for any of the speakers between HL and HH verbs. This can be seen in the following graphs, in which F0 Fall Delay is plotted against Syllable Duration. Datapoints for the HH verbs are plotted with hollow squares, and those for HL verbs are plotted with solid triangles.

Fig.1

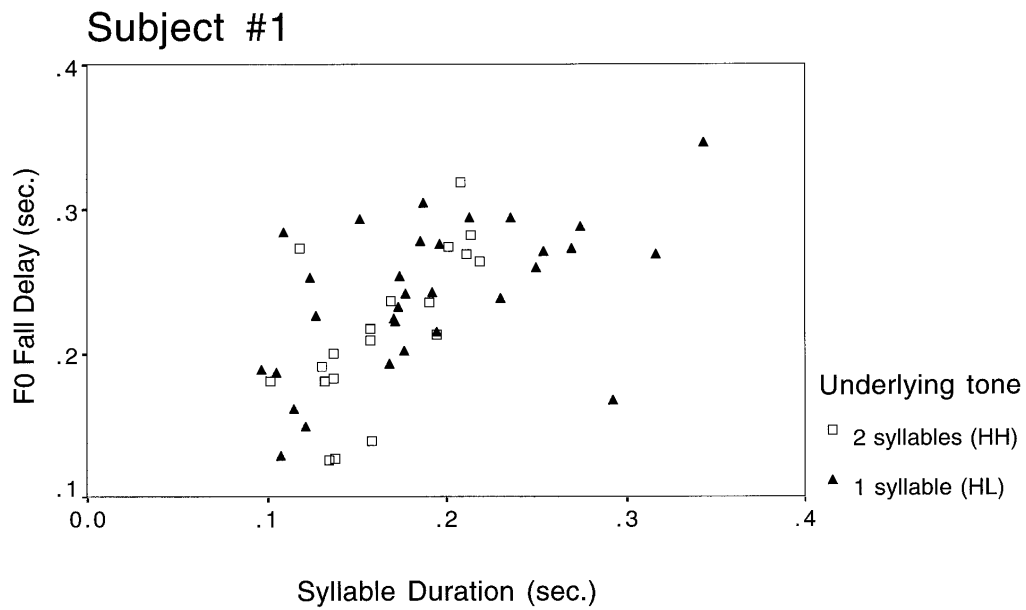


Fig. 2

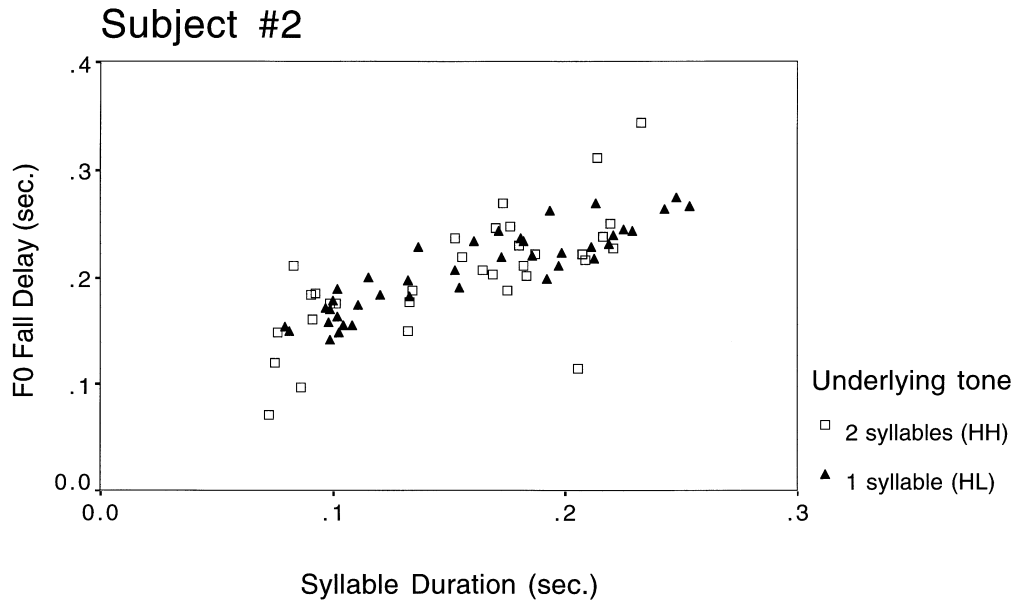


Fig. 3

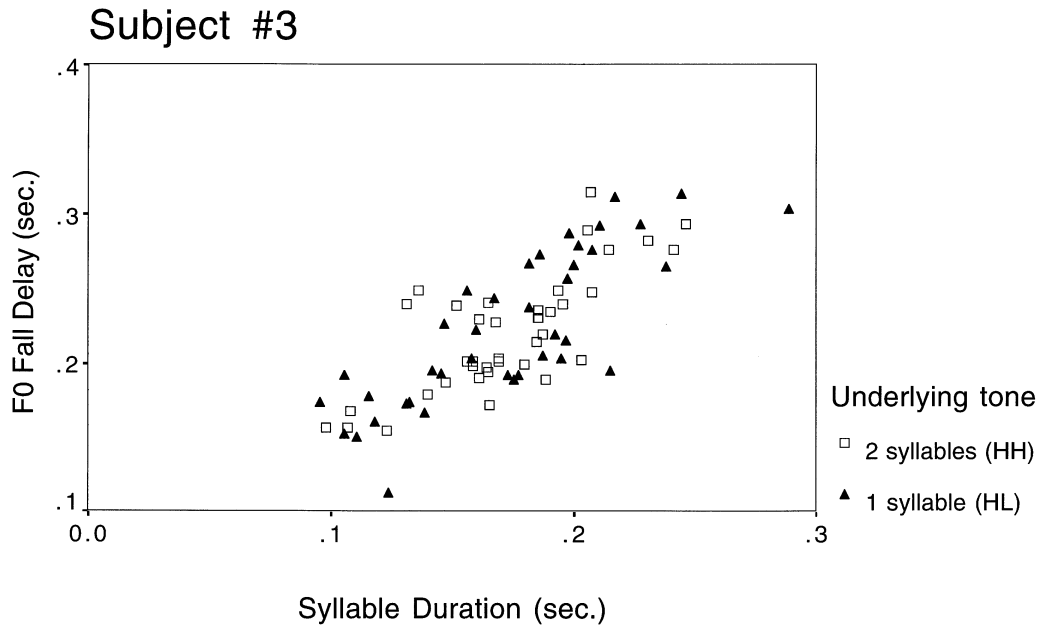


Fig. 4

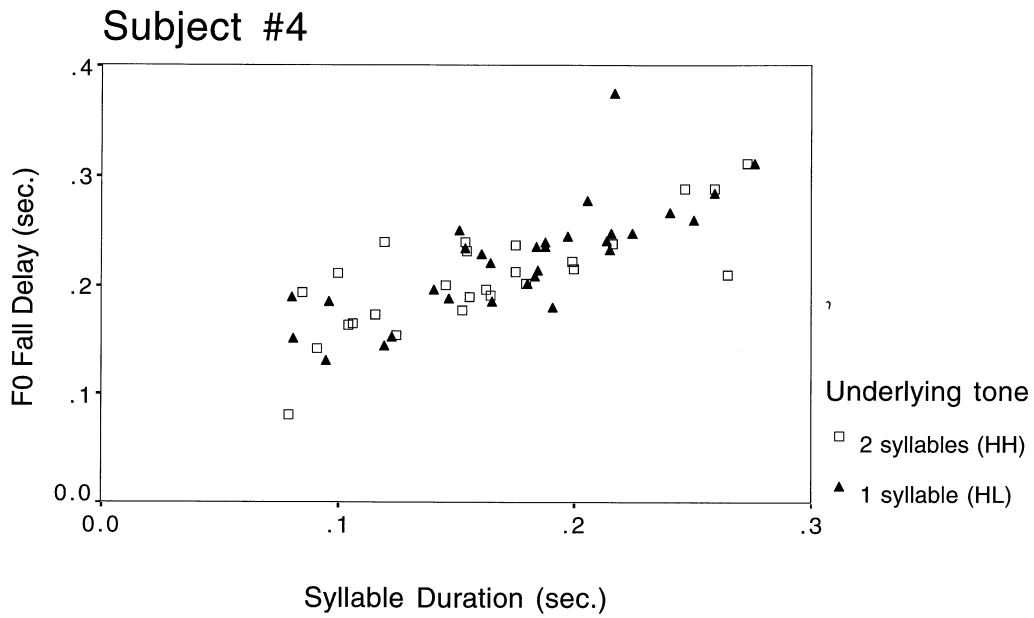
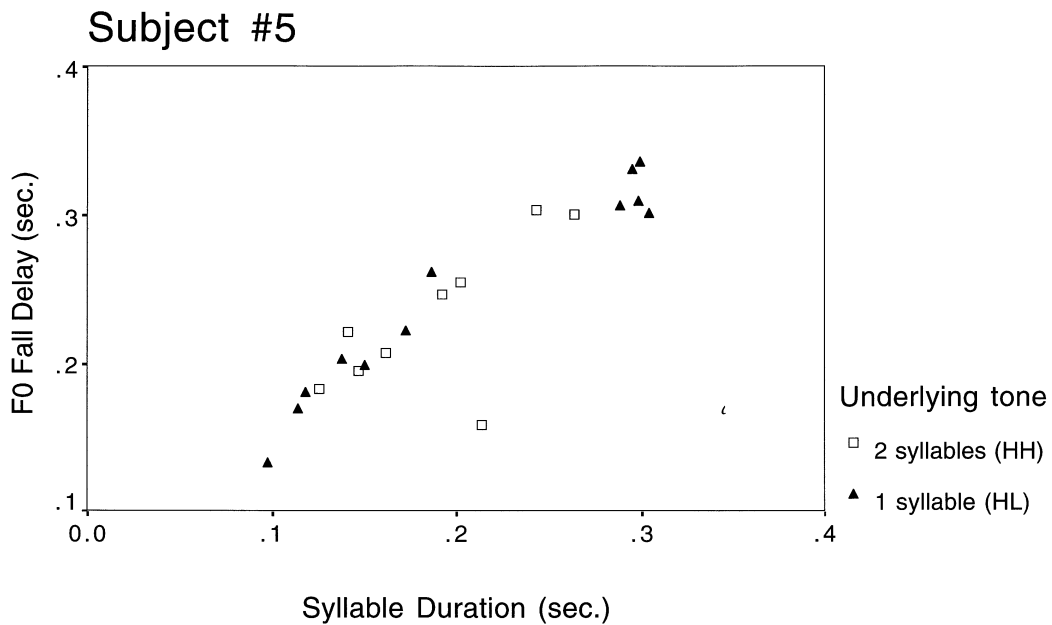


Fig. 5



It can be seen that F0 Fall Delay increases linearly with Syllable Duration. This suggests that the f0 fall comes at a roughly constant proportion of the syllable's duration. It can also be seen that

the squares and triangles overlap throughout their range, suggesting that there is no difference in this proportion according to underlying tone class (HH vs. HL).

Dividing F0 Fall Delay into Syllable Duration for each token, we obtain a measure of the proportion of the syllable duration at which f0 began to fall – a measure we will call *Relative F0 Fall Delay*. The means for this measure for the two underlying tone classes are given in Table 1.

Table 1: Mean Relative F0 Fall Delay by underlying tone class

<i>Subject</i>	<i>1 Syllable /HL/</i>	<i>2 syllables /HH/</i>	<i>T-test</i>
1	1.35	1.34	$t_{48} = 0.09$, n.s.
2	1.36	1.38	$t_{71} = -0.23$, n.s.
3	1.30	1.29	$t_{75} = 0.19$, n.s.
4	1.31	1.34	$t_{55} = -0.43$, n.s.
5	1.26	1.25	$t_{19} = .06$, n.s.

For all speakers, the f0 fall occurred at around 130% of the duration of the first syllable, which puts it within the onset of the following syllable. There was no significant difference between the two tone classes in this measure, as can be seen from the results of independent samples t-tests, reported in the last column of Table 1.

In order to further explore the temporal relation between F0 Fall Delay and Syllable Duration, multiple linear regression analyses were performed in which F0 Fall Delay was the dependent variable, and the predictor variables were Syllable Duration, and Underlying Tone Class (HL was coded as 0, and HH as 1). The results are given in Table 2.

Table 2: Multiple Regression Analysis of F0 Fall Delay

<i>Subject</i>	<i>Coefficient for Syllable Duration</i>	<i>Coefficient for Underlying Tone</i>	<i>Constant</i>	<i>R²</i>
1	.51	-.01	.14	.33
2	.71	-.00	.09	.59
3	.93	-.00	.06	.63
4	.72	-.01	.09	.61
5	.79	-.00	.08	.82

Syllable duration and the constant were significant factors for all speakers, meaning that removing either from the equation would significantly reduce its predictive fit. Underlying tone had not significant effect for any speaker. The R^2 value in the last column represents the proportion of variance in F0 Fall Delay that is accounted for by these simple models. For all speakers this proportion was significantly greater than 0, and, except for Subject #1, the models account for most of the variance in the measure.

As noted above, the f0 fall generally began within the onset of the syllable following the test syllable. This onset varied a great deal in duration, from a brief tap for /r/ to an extended closure for /ŋ/ (written *ny*). Arvaniti, Ladd and Mennen (1998) have shown that the timing of an accentual peak in Modern Greek depends on the whole interval from the beginning of the accented syllable to the onset of the first post-accent vowel, an interval they call *C0-to-V1*. This suggests that in Shona some of the variability in the f0 fall delay might be due to variability in the duration of the onset in which the peak occurs.

To test this hypothesis, multiple regression analyses were performed in which the *C0-to-V1* interval was the predictor of f0 fall delay, instead of Syllable Duration. The results are summarized in Table 3.

Table 3: Multiple Regression Analysis of F0 Fall Delay

<i>Subject</i>	<i>Coefficient for C0-to-V1</i>	<i>Coefficient for Underlying Tone</i>	<i>Constant</i>	<i>R²</i>
1	.55	-.01	.10	.32
2	1.00	-.00	-.02	.73
3	.86	-.00	.02	.55
4	.91	-.00	.12	.56
5	1.27	-.01	-.09	.88

In some cases, these models have a higher R^2 value than the corresponding model in which Syllable Duration is a predictor, while in other cases the Syllable Duration model has a higher R^2 . The absolute values of the residuals were calculated for all models, and paired t-tests were performed to determine whether one model had a significantly better fit to the facts. In only one case was the difference significant: for Subject #2 the model based on C0-to-V1 had a significantly smaller errors than the model based on Syllable Duration ($t_{72} = 2.14$, $p = .04$).

3. Conclusion

We have seen that the f0 distinction in Shona between HH and HL is neutralized before a word beginning with a high tone, as has been proposed on the basis of impressionistic data in the literature on Shona tone cited at the beginning of the paper. The statistical analysis of the quantitative instrumental data presented here demonstrates that the analysis of this phenomenon as a categorical, phonological process is not an artifact of the inherently categorical nature of the transcription data that was available up to now. The retraction is categorical, not gradient.

These speakers of Shona thus offer an example of a categorical counterpart to the gradient tone retraction that has been noted in the literature on f0 timing. As noted above, the f0

peak realizing a high tone has been found to occur significantly earlier in the high-toned syllable when the next syllable is high-toned (Silverman and Pierrehumbert 1990, Prieto, van Santen and Hirschberg 1995). Silverman and Pierrehumbert (1990) find further that the effect extends to cases in which the following high tone is further away than the next syllable, but that the effect diminishes with distance. This extension to non-adjacent elements is paralleled in phonology in the finding that in Arabic roots, pairs of similar consonants are more under-represented the closer they are to each other in the root (Frisch, Broe and Pierrehumbert 1997).

In both phonology and phonetic implementation, then, the juxtaposition of identical high tones can be avoided through retraction of one away from the other. In both domains there must be a version of the Obligatory Contour Principle. In both domains this principle conflicts with a conservative tendency to keep the tones in one particular place. The main difference between the two phenomena lies in how the proximity of the two elements is reckoned. In the phonological pattern it is reckoned in terms of categories such as syllables, while in the phonetic pattern it is in terms of milliseconds. The OCP as a phonological constraint constrains what tone is associated with what syllable (or mora), while the phonetic counterpart constrains where in a syllable an f_0 landmark occurs.

A similar parallelism occurs with regard to phrase-final position. Lexical high tones avoid phrase-final position. In some languages, such as Chichewa, word-final high tones are retracted to the penult if they occur in phrase-final position (Myers 1999). This mimics a phonetic pattern, according to which a high tone is realized earlier in the syllable, the closer it is to the end of a word or phrase (Silverman and Pierrehumbert 1990, Prieto, van Santen and Hirschberg 1995). Again, the retraction pattern in the two domains share the triggering constraint – avoidance of high tone in phrase-final position. They differ in how proximity to the end of the phrase is reckoned: in terms of tone-bearing units, or in terms of real time.

Such close parallelism can be expressed in any model in which the mechanisms that shape the distribution of phonological categories are isomorphic with those that shape

production. Pierrehumbert (1994), for example, has argued that both phonological and phonetic patterns should be expressed in probabilistic models, while Flemming (1997) has argued that phonetic implementation should be expressed in terms of quantitative constraints with an Optimality Theory organization. The correspondence between the two sorts of constraint can be expressed in models in which the constraints on phonology are imported from the phonetics (e.g. Myers 1997b), or in one where a systematic correspondence is enforced between the two (Hayes 1996).

The close correspondence between phonological and phonetic patterns is a problem, on the other hand, for models in which the two domains are given quite different formal structures. This would be the case, for example, if one espoused an Optimality-Theoretic approach to phonology, but maintained the original string-rewrite formalism for phonetic implementation. Likewise, it would be problematic in a derivational model of phonology, since phonological rules in such a model map phonological representations to phonological representations, while phonetic implementation rules map such representations to phonetic realizations.

Bibliography

- Arvaniti, A., D. R. Ladd, and I. Mennen (1998). "Stability of tonal alignment: The case of Greek prenuclear accents". *Journal of Phonetics* 26. 3-25.
- Carter, H. and G. Kahari (1981). *Shona Language Course*. African Studies Program, University of Wisconsin, Madison.
- Cohn, A. (1993). "Nasalization in English: Phonology or phonetics". *Phonology* 10. 43-82.
- Fivaz, D. (1970). *Shona Morphophonemics and Morphosyntax*. University of the Witwatersrand Press, Johannesburg.

- Flemming, E. (1997). "Phonetic optimization: Compromise in speech production". In V. Miglio and B. Morén (eds.) *Selected Phonology Papers from H-OT-97 (University of Maryland Working Papers in Linguistics 5)*. University of Maryland, College Park, 72-91.
- Fortune, G. (1985). *Shona Grammatical Constructions* (3rd ed.). Mercury Press, Harare.
- Frisch, S., M. Broe and J. Pierrehumbert (1997). "Similarity and phonotactics in Arabic". Available from Rutgers Optimality Archive (<http://ruccs.rutgers.edu/roa.html>).
- Hayes, B. (1996). "Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. ROA-158, Rutgers Optimality Archive, <http://ruccs.rutgers.edu/roa.html>.
- Leben, W. (1973). *Suprasegmental Phonology*. PhD dissertation, MIT, Cambridge, published by Garland Press, New York.
- McCarthy, J. (1986). "OCP effects: Gemination and antigemination". *Linguistic Inquiry* 17. 207-263.
- Myers, S. (1990). *Tone and the Structure of Words in Shona*. Garland Press, New York.
- Myers, S. (1997a). "OCP effects in Optimality Theory". *Natural Language and Linguistic Theory* 15. 847-892.
- Myers, S. (1997b). "Expressing phonetic naturalness in phonology". In I. Roca (ed.) *Constraints and Derivations in Phonology*. Oxford University Press, Oxford, 125-152.
- Myers, S. (1998). "Surface underspecification of tone in Chichewa". *Phonology* 15. 367-391.
- Myers, S. (1999). "Tone association and f₀ timing in Chichewa". *Studies in African Linguistics* 28. 215-239.
- Odden, D. (1981). *Problems in Tone Assignment in Shona*. PhD. dissertation, University of Illinois, Urbana-Champaign.
- Pierrehumbert, J. (1994). "Knowledge of variation". *CLS 30: The Parasession on Variation in Linguistic Theory*. Chicago Linguistic Society, Chicago, 232-256.
- Pierrehumbert, J. and M. Beckman (1988). *Japanese Tone Structure*. MIT Press, Cambridge.

Prieto, P., J. van Santen and J. Hirschberg (1995). "Tonal alignment patterns in Spanish".

Journal of Phonetics 23. 429-451.

Prince, A. and P. Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative*

Grammar (Technical Report #2). Rutgers Center for Cognitive Science, Rutgers

University, Piscataway, N.J.

Silverman, K. and J. Pierrehumbert (1990). "The timing of prenuclear high accents in English".

In J. Kingston and M. Beckman (eds.) *Papers in Laboratory Phonology I: Between the*

Grammar and Physics of Speech. Cambridge University Press, Cambridge. 72-106.

Stevick, E. (1969). "Tone in Bantu". *International Journal of American Linguistics* 35. 330-341.