Downstepped Pitch Accent in American English is Categorical and Predictable

Tae-Jin Yoon; Jennifer Cole
University of Illinois at Urbana-Champaign, tyoon@uiuc.edu; jscole@uiuc.edu

Résumé:
A travers une analyse acoustique et une expérience didactique sur machine, cette présentation met en évidence une distinction de catégorie en anglais américain entre les tons hauts accentués H dont le registre est abaissé (downstep: !H*) et ceux dont il ne l'est pas. Cette étude offre une explication quant aux découvertes contradictoires de recherches acoustiques precedents (Liberman et Pierrehumbert 1984; Dainora 2001a,b) sur le downstep qui donnent à réfléchir sur l'hypothèse que le statut de l'accent !H* en anglais américain constituerait une catégorie légitime.*

1. Introduction

The paper presents evidence from acoustic analysis and a machine learning experiment for a categorical distinction between downstepped and non-downstepped high-toned pitch accents (H* vs. !H*) in American English. The present study offers an explanation for the contradictory findings from prior acoustic studies of downstep (Liberman & Pierrehumbert 1984 *vs.* Dainora 2001a,b), which call into question the status of the downstepped accent in American English as a legitimate prosodic category. Dainora (2001a,b) suggests that there is a single phonological High tone that can be used in the specification of pitch accent melody, and "downstepped" pitch accents are illusory, being no more than a subset of variants taken from the normal distribution of H* peak values.

2. Is there !H* in American English?

Pierrehumbert (1980) adopts the analysis of a downstepped pitch accent as a phonologically derived feature. Later work in intonation building on Pierrehumbert's model allows the possibility that downstep is non-automatic and encoded as a contrastive tonal category (e.g., Ladd 1983; Beckman & Ayers 1997). Pierrehumbert (2000) notes, however, that despite the experimental findings of Liberman & Pierrehumbert (1984) in support of downstep, the categorical status of !H* is not substantiated by any large-scale study of naturally occurring speech. More recently, the categorical status of !H* is called into question by Dainora (2001a,b). Based on the analysis of Radio News speech (news stores read by 7 professional announcers), Dainora investigates the status of !H* by comparing the pitch drop in the tonal sequences (H* !H*) and (H* H*). She shows that the pitch drop measure defines a unimodal distribution, where H* and !H* belong to opposite ends of a single distribution in the F0 dimension, in contradiction of earlier claim that !H* forms a distinct tonal category.

3. Categorical Status of !H*

We argue that Dainora's study fails to consider the effects of peak height on the pitch measure; specifically, the F0 peak of the first H* in the sequence might condition the magnitude of the pitch drop to a following peak. In an alternative analysis developed here, we analyze the peak of the second pitch accent (both H* and !H*) in relation to the peak of the preceding H* in the target sequences. We apply this analysis to the same set of data from the Boston Radio News corpus used by Dainora, and on data from a Maptask corpus of spontaneous speech (Shattuck-Hufnagel et al. 2004) produced by 1 female speaker. Using regression analysis methods, we show that H* and !H* form two distinct distributions when the F0 peak is plotted against the peak height of a preceding H*. In regression analysis for peaks in the sequence H*H* in the Boston Radio News corpus, the slope and intercept is 1.0 and 15.93, respectively ($Y=1.0X + 15.93$). For peaks in the sequence H*!H*, the slope and intercept of the regression is 0.5 and 63.95, respectively ($Y=0.5X+63.95$) (see Figure 1). For peaks in sequences of H*H* in the Maptask corpus, the slope and intercept of linear regression is 0.82 and 36.60, respectively ($Y=0.82X+36.60$), and for peaks in the H*!H* sequence, the slope and intercept values are 0.66 and 50.56, respectively ($Y=0.66X+50.56$).

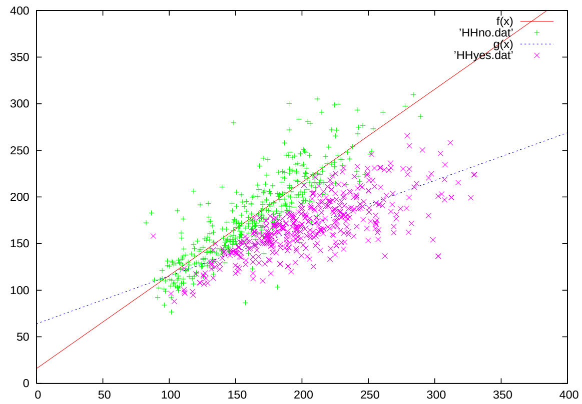

*Figure 1: Scatter plot of H*H* versus H*!H* in the Boston Radio News Corpus.*

We also conduct a machine learning experiment that classifies a pitch accent as H* or !H* based on the F0 peak of the preceding H* accent. The classification experiment achieves between 86.22 % accuracy in a speaker-independent design for the Boston Radio News data and 92.21 % accuracy in a speaker-dependent design for the Maptask data, as in Table 1.

Table 1: Confusion matrix of predicting H* and !H* from Boston Radio News (left) and Maptask (right) Corpora. Observed categories are listed in the columns and predicted categories are listed in the rows. Overall accuracy is 86.22% for Boston Radio News corpus and 92.21% for Maptask corpus.

	Boston Radio News		
	H*	!H*	Precision
H*	380	55	87.36%
!H*	63	358	85.04%
Recall	85.78%	86.68%	

	MapTask		
	H*	!H*	Precision
H*	430	14	96.85%
!H*	48	304	86.36%
Recall	89.96%	94.60%	

4. Discussion and Conclusion

In this paper, we have shown that the downstepped pitch accent (!H*) indeed constitutes a category different from non-downstepped, or normal high pitch accent (H*). Various unknown factors may influence the speech patterns found in natural speech, obscuring the comparison with speech obtained in a laboratory setting. Statistical methods can in some cases be applied to compensate for uncontrolled factors. The experimental findings from naturally occurring speech corpora provide evidence for !H* as a distinct prosodic category, contrary to Dainora (2001a,b) and in support of the findings of Liberman & Pierrehumbert (1984).

References

BECKMAN, M. & AYERS, G. (1994). *Guidelines for ToBI Labeling.* Online MS and accompanying files. Available at http://www.ling.ohio-state.edu/phonetics/E_ToBI.

DAINORA, A. (2001a). Eliminating downstep in prosodic labeling of American English. In Bacchiani *et al*. (eds), *Proceedings of the Workshop on Prosody in Speech Recognition and Understanding*, 41-46.

DAINORA, A. (2001b). *An Empirically based Probabilistic model of Intonation in American English.* Ph.D. dissertation, University of Chicago, IL.

LADD, D.R. (1983). Phonological Features of Intonational Peaks, *Language*, 59, 721-759.

LIBERMAN, M. & PIERREHUMBERT, J. (1984). Intonational Invariance under Changes in Pitch Range and Length. In Aronoff, M.. & Oehrl, R.T. (eds), *Language Sound Structure,* 157-234, Cambrige, MA, MIT Press.

PIERREHUMBERT, J. (1980). *The Phonetics and Phonology of English Intonation.* Ph.D. dissertation, MIT.

PIERREHUMBERT, J. (2000). Tonal Elements and Their Alignment. In Horne, M. (ed), *Prosody: Theory and Experiment,* 11-36, Dordrecht, Kluwer.

SHATTUCK-HUFNAGEL, S., DILLEY, L., VEILLEUX, N., BRUGOS, A. AND SPEER, R. (2004). F0 peaks and valleys aligned with non-prominent syllables can influence perceived prominence in adjacent syllables. In *Proceedings of the ICSA International Conference on Speech Prosody 2004,* 705-708, Nara, Japan.