ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2012
Sixth International Conference
Shanghai, China
May 22-25, 2012

Dialect Adaptation and Two Dimensions of Tune

James Sneed German

Division of Linguistics and Multilingual Studies, Nanyang Technological University, Singapore

jsgerman@ntu.edu.sg

Abstract

Adaptation to an unfamiliar dialect of one's native language presents a special case for prosodic learning, since most other aspects of the grammar are held constant. This study explores the representation of two dimensions of tune through a series of experimental tasks in which speakers of American English attempt to directly imitate and then generalize the dialectal features of a native speaker of Glasgow English. The results show that speakers are able to modify both f0 peak timing and f0 excursion in order to approximate the target dialect, and that they do so both during direct imitation and when generalizing to new sentences. The findings suggest that peak timing and excursion are not only represented differently, but that learning progresses differently for the two dimensions in going from direct imitation to generalization.

Index Terms: dialect learning, intonation, peak alignment

10.21437/SpeechProsody.2012-109

1. Introduction

When a new language is introduced into a community of adult speakers, its prosodic characteristics are generally not replicated precisely, but may diverge from the source language in various ways due to, for example, learning by approximation with L1 categories [1, 2], constraints on perception imposed by L1 [3], and differences in the input distributions. It is important to note that such effects may operate at different levels. In the case of intonation patterns, for example, learning takes place at at least two levels: Speakers not only acquire an inventory of tone units and rules about how to combine and sequence them, but they also must acquire a set of phonetic implementation rules that specify how the underlying tone sequences should be pronounced [4]. Thus, the various types of "interference" mentioned above may take place at either level. Speakers may reinterpret intonation patterns in terms of preexisting tonal inventories in L1, as has been suggested by [5] for Hong Kong English. Alternatively, they may replicate certain aspects of the L2 tone inventory, but may rely on L1 phonetic implementation rules for their realization [6, 7].

Finally, the distinction may emerge during learning itself. According to one group of theories, experiences of individual speech events are stored in memory with rich phonetic detail as well as various kinds of indexical information concerning category membership, and the lexical and social context ([8], [9]). In learning, abstract coding categories emerge as large numbers of such stored experiences give rise to robust statistical generalizations [10]. Crucially, however, this process may be affected by the category labels that are initially assigned at the time the speech event is experienced and recorded. Learners, in other words, may perceive and remember an L2 speech event perfectly, yet that same event may actually contribute to a "non-native" bias during retrieval, due to the fact that it was originally coded according to a non-native pattern. In order to understand how differences between prosodic systems affect learning and change, therefore, it is crucial to consider not just how native L2 patterns are perceived and imitated, but also to compare these to the patterns that emerge during generalization. In other words, given that speakers can modify their intonation patterns to resemble those of another system, how do these modifications differ when the learner has access to a native-like phonetic representation, versus when the learner must implement the new model from the ground up?

This study is a first attempt to address these issues by comparing how speakers of American English learn and generalize two dimensions of tune during controlled exposure to an unfamiliar dialect (Glasgow English). Since related dialects may be very similar in terms of lexicon, syntax, and segmental patterns, dialect learning is an attractive tool for the study of prosody. For one thing, the high degree of overlap frees up working memory so that learning targets those features of most interest to the researcher. Additionally, it provides learners with ready access to rich semantic and contextual information that further facilitates memory and allows them to build realistic and holistic representations.

The two dimensions of tune addressed by this study are f0 peak timing and f0 excursion. These features are explored in connection with a specific intonational contour of the Glasgow English dialect, namely the *rise-fall*, or **L*H H-L%** in the GlaToBI transcription system [11]. As Figure 1 illustrates, that contour is characterized by a relatively long rise from a pitch trough early in the nuclear syllable to a pitch peak that occurs late in the syllable. The peak is then followed by a sharp drop in f0, which remains low through the rest of the phrase.

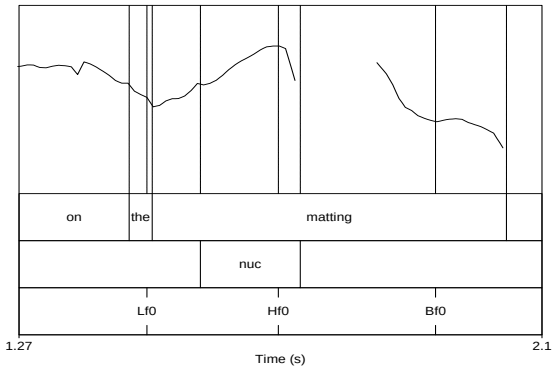

Figure 1: *Typical sentence-final rise-fall pattern (L*H H-L%) of Glasgow English.*

As illustrated in Figure 2, in a similar context (sentence-final trochee in a neutral focus declarative sentence), American English speakers typically produce a falling contour characterized by a moderately high peak relatively early in the nuclear syllable (H* L-L% in ToBI [12]). In addition to having a later peak then, the Glasgow English pattern typically involves a much larger f0 excursion.

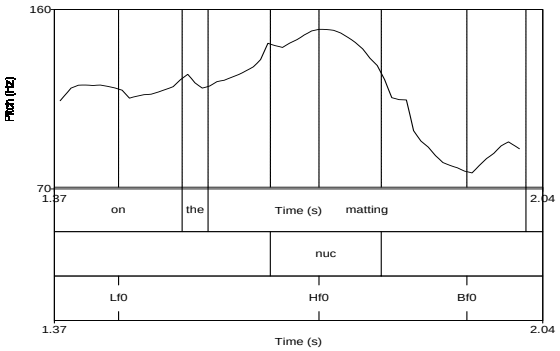

Figure 2: *Typical sentence-final high falling pattern (H* L-L%) of American English*

Both dimensions of phonetic variation are demonstrably relevant for the interpretation of tune, though they differ in terms of what is understood about their specific role in this regard. Specifically, there is considerable evidence showing that f0 peak timing gives rise to robust, language-specific categorical distinctions [13, 14, 15], whereas the status of pitch excursion in this regard enjoys much less of a consensus. An early view was that for English, the pitch continuum is divided into as many as four pitch levels [16]. [17] and subsequently [18] proposed a simplification of the pitch level approach, whereby only two abstract pitch levels, high and low, are relevant for tune, and most of the remaining variation associated with f0 can be reliably modeled in terms of scalar variables such as prominence, declination and range. [19] showed that for Dutch, the perception of prominence of the second of two high pitch accents is sensitive to the specific height of the first accent, while more recently, [20] showed that in English, listeners make a category-like distinction when sorting graded excursions into those describing "everyday" versus "unusual" occurrences. As yet, however, there is no consensus regarding the presence or number of categories associated with distinctions in the height of f0 rises and falls.

This study exploits these potential differences in order to explore the issue of how intonational tunes are learned and represented. Specifically, it compares the extent to which learning progresses differently along the two dimensions when speakers try to reproduce tunes in an unfamiliar dialect. More importantly, by comparing both direct imitation and generalization, this study evaluates learning in two different task paradigms that are predicted to involve different capacities of the learner. Since f0 peak timing is categorical, learners should be able to relate D2 patterns to D1 categories, and learning should exhibit a relatively high degree of stability in going from imitation to generalization. By contrast, if f0 excursion is gradient, then even "successful" learning should exhibit a higher degree of variability.

2. Methods

2.1. Participants

Two male and two female Northwestern University undergraduate students participated in this study for course credit. All were age 20 at the time of the study, had lived in the U.S. since birth, and reported having English as their first and dominant language.

2.2. Materials

The speech data used in this study were collected as part of a larger study investigating the ability of speakers to learn and generalize allophonic realization patterns of an unfamiliar dialect [21]. Specifically, the materials were designed to test whether speakers are able to imitate, learn and generalize the Glasgow English pattern by (i) reassigning the [t^h] allophone of /t/ to falling stress environments where it would normally appear as [ɾ], and (ii) reassigning the [ɾ] allophone of /t/ to lexically specified occurrences of the phoneme /r/.

The total set of experimental materials consisted of four printed sets of forty-eight sentences of English plus recordings of those sentences as produced by a native speaker of Glasgow English. The sentences contained no lexical items specific to Glasgow English, so the Glaswegian recordings amounted to exemplars of "Glasgow-accented English" rather than the Glasgow dialect of English per se. These recordings were made in a sound-attenuated room using a Shure SM81 microphone and digitally encoded using ProTools 2.0 audio software at a sampling rate of 55 kHz. They were resampled at 22.5 kHz before being recorded onto audio CD for playback.

2.3. Procedure

During the experiment, participants' speech was recorded during four separate tasks. In the Baseline task, participants read sentences from one of the four scripts using their native dialect (i.e., Standard American English). This task was completed first, so the participants had not yet been exposed to the Glasgow dialect. In the Training 1 task, participants were told that they would be listening to a recording of a speaker using an "unfamiliar dialect" and that they should try to imitate the way he said each sentence. They were not given any information about the origin of the dialect, and they were given no instructions regarding specific characteristics of the dialect that they should emphasize. Participants would listen to the Glasgow speaker producing each sentence in the set while following along on the printed script, and then imitate the sentence into the microphone. In the Training 2 task, this procedure was repeated once using exactly the same materials immediately after its first iteration. Finally, in the Generalization task, participants were given a third set of sentences, which they had not previously seen nor heard the Glaswegian speaker produce, and were asked to continue imitating the accent without the aid of any recordings. All tasks were recorded digitally according to the method described in 2.2 for the Glasgow English speaker.

The present study addresses learning and generalization of a specific intonational pattern in an unfamiliar dialect. Since the materials described above were designed to address allophonic learning, the segmental characteristics of the materials (e.g., word length, syllable structure, voicing) are not ideally controlled for intonational analysis. Moreover, the Glasgow speaker produced a variety of intonational patterns, which the participants were subsequently exposed to in the course of the Training tasks. So while the **L*H H-L%** tune was the most common choice for participants in the Generalization task, it was not possible to insure that they would produce this tune for every item in that task. For these reasons, the data used in the present study consist of a subset of the recordings described above. Specifically, six productions from each participant were selected from each of the Baseline, Training 2 and Generalization tasks.

For all tasks, items were limited to those in which the segmental characteristics of the final two syllables constitute a trochaic pattern. For the Baseline task, sentences were further restricted to those exemplifying a typical sentence-final declarative contour characterized by **L- H* L-L%** in American English ToBI. This step was taken to insure a degree of comparability for target contours in both the Glasgow English speech and the attempted imitations.

For the Training tasks, items were chosen based on whether the Glasgow speaker used the **L*H H-L%** pattern. A similar criterion was used for selection of items in the Generalization task, though in that case, the Glasgow speaker's productions were not relevant, so tokens were chosen based on the tune chosen by the participant. Together, these steps ensured that a consistent and comparable set of intonational targets was being measured and evaluated across the Baseline, Training and Generalization tasks.

2.4. Analysis

Two measurements were made based on the sentence-final f0 contour of each token. Peak alignment was measured as the temporal distance between the f0 peak of the nuclear rise and the end of the nuclear syllable. The duration of the rhyme was also measured, and proportional peak delay was calculated as the ratio of peak delay to rhyme duration [14]. This step was taken in order to adjust for differences in speaking rate. F0 excursion was measured as the difference between the f0 peak of the final nuclear rise and the trough or elbow immediately preceding it. For the American English pattern, this generally coincided with the end of the **L-** preceding the nuclear contour, while for the Glasgow English pattern, this point corresponded to the onset of the **L*H** nuclear accent itself, which in turn generally coincided with the onset of the nuclear syllable as is typical for the **L*H** accent [11]. To partially adjust for both speaker- and utterance-specific differences in pitch range, f0 measurements were converted to the Bark scale prior to subsequent analysis.

3. Results

Figure 3 shows the means for peak delay by speaker and by task, and compares them to the means for the Glasgow English speaker's productions from the corresponding training task materials (Glasgow targets). The results of a two-way ANOVA for independent samples confirm that the means differed across the three tasks ($F(2, 55)=20.22$, $p<0.0001$) though not by speaker ($F(3, 55)=0.81$, $p=0.49$). Overall, f0 peaks occurred much earlier in the Baseline task than in the other tasks, and this was confirmed in a post-hoc Tukey HSD comparison ($p<0.01$), though the same comparison revealed no significant difference between the Training and Generalization tasks. The Glasgow speaker produced peak delays that were close to, but partially retracted from, the end of the nuclear syllable, and the mean peak delays in the Training and Generalization tasks generally reflect this pattern. Interestingly, the mean peak delays in the Training and Generalization tasks appear to be shorter overall (i.e., less negative) than those for the corresponding Glasgow English targets. These differences were found to be significant in two independent samples two-way ANOVAs (vs. Training: $F(1, 40)=6.42$, $p<0.05$; vs. Generalization: $F(1, 40)=4.74$, $p<0.05$). This suggests that while speakers were successful at modifying their peak delay patterns *in the direction of* the Glasgow English pattern, they were not necessarily successful in matching the pattern in terms of approximating its mean and distribution.

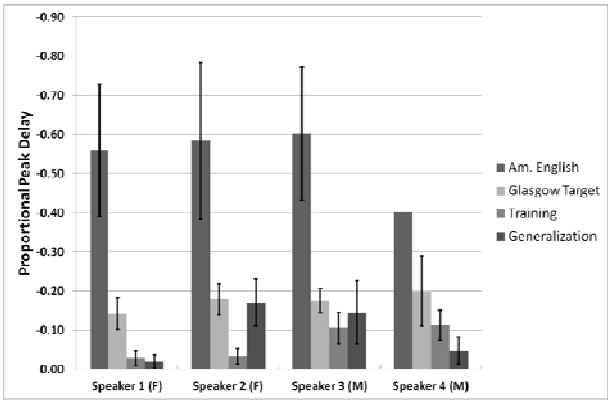

Figure 3: *Peak delay as a proportion of rhyme duration by task for four speakers. Whiskers represent the standard error of the mean.*

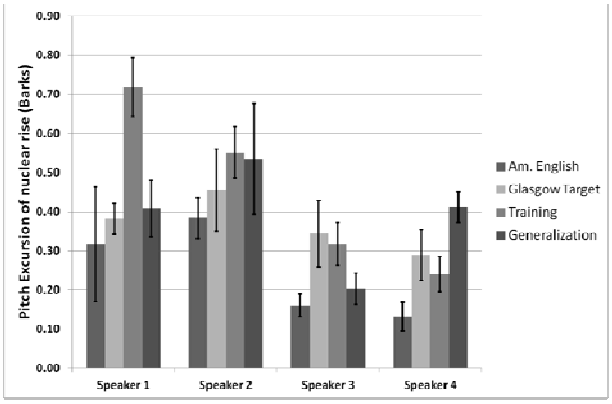

Figure 4: *Excursion in Barks by task for four speakers. Whiskers represent the standard error of the mean.*

Figure (4) shows the means for excursion by speaker and by task and compares them to the means for the corresponding Glasgow targets. Overall, the Glasgow English excursions were larger than the Baseline American English excursions. In an independent samples two-way ANOVA, there were main effects of both task ($F(2, 59)=9.69$, $p<0.0001$) and speaker ($F(3, 59)=7.06$, $p<0.005$) with no significant interaction. For all four speakers, the mean excursion for the Baseline task was lower than for either the Training or Generalization tasks; this was confirmed by a post-hoc Tukey HSD comparison ($p<0.05$). Although the Training task means were higher than the Generalization task for three out of four speakers, the same post-hoc comparison revealed no significant difference between those two tasks. All four subjects therefore were able to successfully augment their pitch excursions in the Training and Generalization tasks, and for the most part did so in a way that closely matched the Glasgow English pattern. Speaker 1, however, produced pitch excursions in the Training task that greatly exceeded those of the Glasgow English targets. A comparison of the Glaswegian targets and Training task revealed a marginally significant difference ($F(1, 40)=3.42$, $p=0.071$). It is likely that this marginal effect was largely driven by Speaker 1's productions. A comparison of the Glasgow targets and the Generalization task revealed no significant difference ($F(1, 40)=0.25$, $p=0.62$).

4. Discussion

The above results establish that speakers are able to modify both their f0 peak timing and the f0 excursion associated with a nuclear contour when trying to approximate the pattern of an unfamiliar dialect. Moreover, they did this both when attempting a direct imitation as well as when attempting to generalize the dialect to new lexical material.

Several interesting patterns emerged. First, speakers were highly successful at modifying their peak timing, though they did not do so in a way that precisely matched the target pattern. This suggests that learning of that particular dimension may have been mediated by either an abstract tonal pattern in the speakers' native dialect, by phonetic implementation rules in the native dialect, or by both. For example, speakers may have been accessing their knowledge of the **L+H*** contour of American English, as well as its implementation, in order to approximate the **L*H** rise in the Glasgow English targets. Unfortunately, there were very few instances of **L+H*** in the larger set of Baseline productions, so further research is needed to test this conclusively.

Interestingly, several of the speakers sounded distinctly non-American in their approximations. An alternative hypothesis then, is that speakers analyzed the Glasgow English contours according to an abstract tonal sequence in their native inventory, were able to modify their implementation rules for those sequences, but did so imperfectly. The fact that the distributions in the Training and Generalization tasks cluster near the end of the syllable is noteworthy in this regard. It suggests, among other things, that to the extent that speakers are learning new implementation rules, they may be relying on salient segmental landmarks rather than learning a specific timing pattern per se.

Speakers' approximations of f0 excursion did not differ statistically from the target patterns. However, the high degree of variability for all excursion results, along with the comparatively smaller difference between the Baseline and Glasgow Target patterns, suggests that more data is needed to determine whether speakers are actually able to match the mean and distribution of a learned non-native pattern.

Overall, speakers appear to perform similarly for both direct imitation and generalization. This effect is particularly robust for the peak timing results, and lends further support to the notion that learning is being anchored by native representations. Although statistically, the excursion results are similar in this regard, the trend for three out of four speakers was for the mean in the Generalization task to be somewhat intermediate to that of the Baseline and Training tasks. Furthermore, for one speaker in the Training task, and for one speaker in the Generalization task, the excursions were substantially higher than for either the American English pattern or for the Glasgow English pattern. Together, these two facts suggest a possible difference in how excursion and peak timing are represented. Specifically, it suggests that speakers were able to learn and remember that excursions are generally *larger* for the new dialect, but they were not able to represent the magnitude of this difference. In other words, the results tend to support the conclusion that peak timing, but not f0 excursion, behaves categorically during dialect learning.

5. Conclusions

The results of this study are suggestive of a new paradigm for exploring how intonation patterns are perceived, represented and implemented, as well as what consequences those factors may have for adaptation by adult speakers. It suggests that opportunities for improvement of the approach include larger data sets providing greater statistical power, particularly with regard to type II error, as well as more extensive characterization of D1 characteristics. As this study was drawn from a larger dataset, future research will explore how success in the prosodic domain is correlated with adaptation in other domains, including allophony patterns and vowel quality.

6. References

[1] Best, C. T., McRoberts, G. W., & Goodell, E., "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system", *J. of the Acoustical Society of America,109(2)*, 775-794, 2001.

[2] Flege, J.E., "Second language speech learning: Theory, findings, and problems", in W. Strange [Ed], *Speech perception and linguistic experience: Issues in cross-linguistic research*, 233-277, York Press, 1995.

[3] Werker, J. F., & Tees, R. C., "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life", *Infant Behavior and Development*, *7*(1), 49-63, 1984.

[4] Beckman, M. & Pierrehumbert, J., "Intonational structure in Japanese and English" *Phonology Yearbook III*, 15-70, 1986.

[5] Luke, K.K., "Phonological re-interpretation: The Assignment of Cantonese Tones to English Words", presented at *the 9th International Conference of Chinese Linguistics*, National University of Singapore, 2000.

[6] Atterer, M. & D. R. Ladd. 2004. On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics, 32(2)*, 177-197.

[7] Mennen, I. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, *32(4)*, 543-563.

[8] Goldinger, S. D., "Echoes of echoes? An episodic theory of lexical access", *Psychological Review, 105*, 251-279, 1998.

[9] Johnson, K., "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology", *Journal of Phonetics, 34*, 485-499, 2006.

[10] Pierrehumbert, J. B., "Probabilistic phonology: Discrimination and robustness", in R. Bod, J. Hay, & S. Jannedy [Eds], *Probabilistic linguistics*, 177-228, MIT Press, 2003.

[11] Mayo, C., *Prosodic Transcription of Glasgow English: an evaluation study of GlaToBi, Speech and Language Processing*, masters thesis, University of Edinburgh, 1996.

[12] Silverman, K., Beckman, M. Pitrelli, J. Ostendorf, M., Pierrehumbert, J., Hirschberg, J. & Price, P., "ToBI: A Standard Scheme for Labeling Prosody", *ICSLP 1992*, 867-870, 1992.

[13] Pierrehumbert, J. B., and Steele, S. A., "Categories of tonal alignmnent in English", *Phonetica* 46:181-196, 1989.

[14] Silverman, K. and J. Pierrehumbert, "The Timing of Prenuclear High Accents in English", *Papers in Laboratory Phonology I*, 72-106, Cambridge University Press, 1990.

[15] Redi, L., "Categorical effects in the production of pitch contours in English", presented at 15th International Congress of the Phonetic Sciences, Barcelona, 2003.

[16] Pike, K., *The intonation system of English*, University of Michigan Press, 1945.

[17] Pierrehumbert, J., *The phonology and phonetics of English intonation*, Ph.D. Thesis, MIT, 1980.

[18] Liberman, M. & J. Pierrehumbert, "Intonational Invariance under Changes in Pitch Range and Length", in M. Aronoff and R. Oehrle [Eds], *Language Sound Structure*, 157-233, MIT Press, 1984.

[19] Gussenhoven, C., and Rietveld, A. C. M., "Fundamental Frequency Declination in Dutch: Testing Three Hypotheses", *Journal of Phonetics 16*:355-369, 1988.

[20] Ladd, D. Robert, and Morton, Rachel, "The Perception of Intonational Emphasis: Continuous or Categorical?", *Journal of Phonetics 25*:313-342, 1997.

[21] German, J., Carlson, K. and Pierrehumbert, J., "Reassignment of the flap allophone in rapid dialect adaptation", in progress.