Phonetic Detail and the Role of Exposure in Dialect Imitation

Mariapaola D'Imperio[1,2] *& James Sneed German*[1]

[1]Aix-Marseille Université, CNRS, LPL UMR 7039, 13100, Aix-en-Provence, France
[2]Institut Universitaire de France, Paris, France

james.german@lpl-aix.fr; mariapaola.dimperio@lpl-aix.fr

ABSTRACT

Speakers are able to adjust their prosodic patterns to approximate those of a different dialect, at least when the dialects involved are phonologically similar [6, 7]. Our study explores imitation across two dialects of English (Singaporean and American) whose prosodic systems are phonologically very distinct. Singaporean speakers were recorded both in their native dialect and while attempting to imitate sentences produced by an American English speaker. Our results show that in spite of the structural differences, speakers of Singapore English are able to rapidly adapt and shift from an edge-based system to an accentual system within the time of the experiment, as well as to finely tune the phonetic detail of their intonation patterns in a way that closely tracked that of the American English model speaker. We further show that the degree of variability in successfully reproducing the target values is dependent on amount of exposure to the non-native dialect.

Keywords: dialect imitation; intonation; Singapore English; convergence.

1. INTRODUCTION

Interacting speakers align their phonetic representations to that of their interlocutor, and this unconscious convergence can be seen as a process of spontaneous imitation manifested both at the segmental and the prosodic level ([8, 10], inter alia). As for prosodic convergence, previous studies have mainly reported convergence in global phonetic features, such as pitch range and speech rate, even in the absence of direct social interaction [2].

However, studies on both spontaneous and direct intonational convergence (i.e., dialect imitation) have presented mixed results regarding speakers' ability to accurately imitate phonetic detail of a specific prosodic event. On one side, it has been argued that speakers are only able to reproduce "structural" and "feature" patterns of the prosodic form produced by the interlocutor. In [5], for example, speakers did not accurately reproduce phonetic implementation of intonation phrase boundaries at the duration level. Speakers in that study were, however, able to reproduce structural features, such as pitch accent and tonal boundary placement, as well as pitch accent categories, with a certain degree of accuracy. [3], in line with [5]'s findings, report that, when asked to reproduce previously heard, randomly synthesized intonational contours, British English speakers could only reproduce legal contours for their own varieties, which acted as "attractors" for production according to the authors.

Intonational convergence has also been found in cross-dialect imitation of tonal alignment and pitch scaling features. Note that tonal alignment has both a phonetic and a phonological dimension, given that language varieties can differ in alignment for the same pitch accent category (cf. [1]). In a direct dialect imitation study, American English speakers were asked to imitate an unfamiliar dialect, i.e. Glasgow English, after producing baseline intonation patterns [7]. Tonal alignment measures for rising accents (L+H* and L*+H) showed that American English speakers were able to shift the alignment of their peaks in imitating the late Glasgow peaks, and that they could successfully produce the patterns in a subsequent generalization phase. However, actual phonetic implementation of H peak alignment did not accurately match either the delay or the variability of the Glasgow speaker, and appeared instead to be stably aligned to a different syllable-internal position. [7] suggests that the imitation process may have been mediated by an abstract tonal pattern in the native language and/or by native implementation rules.

A more recent cross-dialectal study on two varieties of Italian [6] showed that both coarse-grained, phonological features of tonal composition of the contour as well as fine phonetic detail of tonal alignment and scaling can be successfully reproduced in an imitation task. However, imitation of the target alignment values was different for Bari speakers imitating Neapolitan speakers than for the other way around, given that the two dialects differ in the number of contrasting LH accents in their inventory (one or two). In fact, Neapolitan speakers (possessing a L+H* and a L*+H contrast), when shifting tonal alignment earlier to reproduce the L+H* of Bari Italian nuclear accents, showed a marked overshoot relative to the model speaker. This suggests that although phonetic detail can be

reproduced to a certain extent, the degree of precision might vary according to the constraints of the native tonal system. Additionally, the degree of exposure to the unfamiliar dialect may explain variability in the accuracy of imitating gradient phonetic detail in alignment.

In this study, we test intonational convergence in a direct imitation task involving two varieties of English that are typologically different in terms of their prosodic and intonational phonology, i.e. Singapore and American English. While Singapore English (SgE) is an edge-based language [9] in which tonal events are timed to occur with the edges of prosodic domains larger than the prosodic word [4], American English (AmE) is a head-based, stress-accent language, such that some tonal events are timed to occur at specific time intervals within a given domain, such as a vowel nucleus. In order for speakers of SgE to adjust to the AmE pattern therefore, they must learn to attend to a very different set of cues when selecting the implementation for f0.

Our main hypothesis is that strong typological differences in the prosodic organization of two dialects will interfere with speakers' ability to imitate across the dialects, especially given that this process requires phonetic, and not-just phonological convergence. We also hypothesize that the degree of exposure to the non-native dialect will affect the degree of accuracy in phonetic implementation of non-native tonal alignment measures.

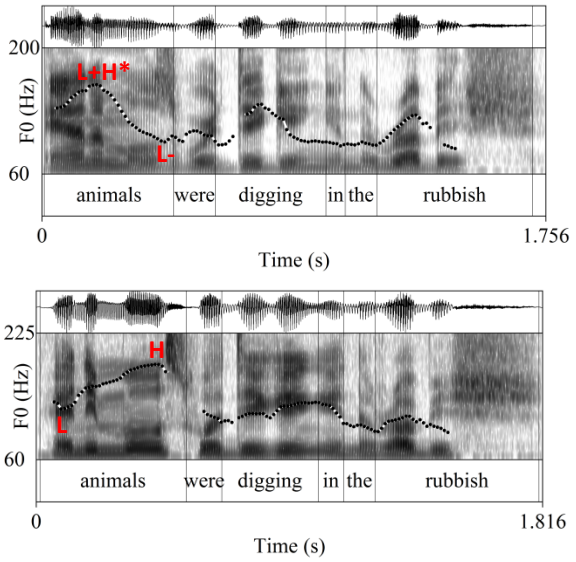
2. METHODS

Participants read sentences out loud under two different task conditions: In their native dialect in the absence of audio prompts (Baseline), and in response to auditory prompts which they were asked to imitate (Imitation).

2.1. Materials

Twelve trisyllabic target words with initial stress and open syllables (e.g., *animals*) were elicited in each of the task conditions. These occurred at the beginning of the sentences in which they appeared, and were followed by sequences designed to elicit consistent phrasing patterns in SgE (e.g., '[*Animals*][*were digging*][*in the rubbish*]'). Target words were further selected on the basis of their lexical frequency (high vs. low) and high familiarity rating [13]. Recordings of the target sentences were taken from a native speaker of AmE, and were produced with a declarative intonation pattern including a L+H* L- pattern on both the target word and the immediately following phrase (Figure 1, top). The SgE pattern for the same target involves a rise from a low f0 (L) at the beginning to a high peak (H) close to the end of the word (Figure 1, bottom). Approximating the AmE pattern therefore requires, among the other things, that SgE speakers adjust the alignment of the first f0 peak to a much earlier position. 24 additional sentences (12 yes/no questions, 12 conditionals) targeting other contours were included, but not analysed for this study.

Figure 1: F0 traces for one target sentence as produced by a speaker of American English (top) and a speaker of Singapore English (bottom).



2.2. Participants

9 male speakers of SgE age 20 to 27 participated in the study. All were students at Nanyang Technological University at the time of the study. SgE prosody differs substantially among ethnic groups [12], so only ethnically Chinese speakers were selected. All were multilingual in at least Mandarin and English.

2.3. Procedure

Prior to the experiment, participants filled out an anonymous questionnaire asking them to estimate the number of hours per week they were exposed to American varieties of English at the time of the experiment, in terms of media (tv, movies, online videos), instruction (lecturers/professors) and friends or acquaintances. For the experimental tasks, participants were seated in front of a computer screen. In the Baseline task, sentences appeared one at a time, and participants were instructed to read them aloud in a natural and conversational style. In the Imitation task, the text of each sentence appeared one second before playback of the corresponding

recording. Participants were instructed to "imitate the way the speaker says the sentence as closely as possible." They were told that the speaker uses a variety of English different from their own, though no mention was made of specific features to be imitated. Participants were allowed multiple attempts at imitation of each trial, though repeated playback of the recording was not allowed. Trials in all tasks were self-paced. After all items had appeared once in the Imitation task, the entire task was repeated once in a different pseudorandom order. Only productions from the Baseline and the second round of Imitation were used for analysis.

2.4. Analysis

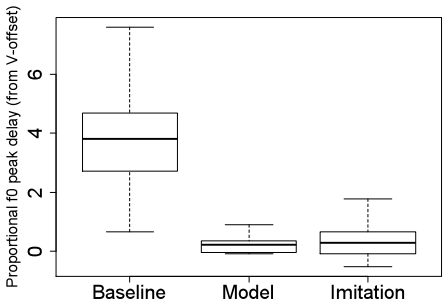
Collected recordings of the participants and the Model Speaker were manually labelled in Praat for (i) the location of the f0 peak in the target word, (ii) the right edge of the target word, and (iii) the edges of the vowel nucleus of the first syllable of the target word. F0 peak delay from the end of the (first syllable) vowel nucleus and the duration of the nucleus were extracted automatically. Proportional Peak Delay (PPD) was calculated by dividing the f0 peak delay by the duration of the nucleus. Since global statistics can obscure how participants deal with variability in the stimuli, we also considered how the imitations compared to the model speaker on a trial-by-trial basis. Specifically, difference scores were calculated by subtracting the PPD of the Model Speaker for that item from the participant's PPD.

3. RESULTS

3.1. Adjustment to f0 peak delay

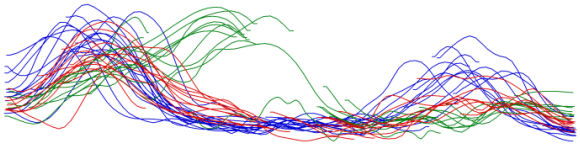
As expected, productions in the Baseline task had peak delays that were close to the end of the target word, and therefore substantially later than productions by the model speaker for the same words. Nevertheless, in the Imitation task, participants were able to implement earlier timing for their f0 peak targets thereby closely approximating the PPDs of the model speaker. This shift can be clearly seen in the boxplot in Figure 2. The effect of task on PPD was modelled using linear mixed effects analysis with subject and item as random intercepts. Unsurprisingly, including task type (Baseline vs. Imitation) as a fixed factor significantly improved the fit of the model relative to a base model with only random intercepts ($\chi^2(1) = 274.16$, $p < 0.001$). Including lexical frequency (high vs. low), however, did not ($\chi^2(2) = 0.28$, $p = 0.60$).

Figure 2: Boxplot of f0 peak delay as a proportion of vowel nucleus duration for SgE participants in two task types and the model AmE speaker.



Peak delay is only one measure of similarity between pitch accents, and cannot necessarily reveal whether speakers were approximating the holistic shape of the model speaker's nuclear contour. Evidence from contour plots like that in Figure 3, however, suggests that participants were, in fact, matching the contours of the target speakers at a very fine degree of phonetic detail that included peak alignment, shape, and scaling of the f0 contour.

Figure 3: F0 contour plots over the target region and following phrase for the model AmE speaker (blue) and one SgE speaker (ID102) in the Baseline (green) and Imitation (red) tasks.

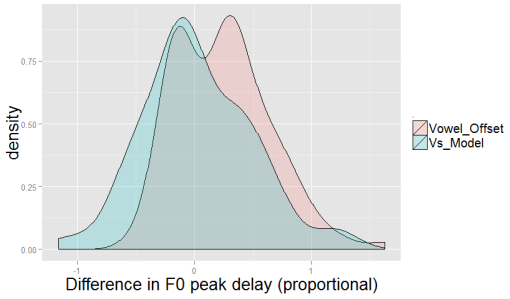


3.2. Difference scores

Within the Imitation task, peak delay values were small in magnitude, suggesting that most of the f0 peaks lay close to the offset of the vowel nucleus. In light of findings by [7] that imitators were making use of segmental landmarks, our results raise the question of whether "successful" imitation in our study was due to close matching of the model speaker's details, or whether this was an artefact of the model speaker's own tendency to align peaks close to the nucleus offset. One way to address this is by comparing raw PPD scores to PPD difference scores. If the imitators are simply aligning to the nucleus offset, then raw PPD should appear more stable and smaller in magnitude compared to the PPD difference scores. If the imitators are approximating the characteristics of the model speaker's specific tokens, however, then the PPD should appear more *un*stable and larger in magnitude compared to the difference scores, since the imitators' peaks should move towards and away

from the nucleus offset accord to what the model speaker is doing on any given trial. As Figure 4 shows, there are in fact, some important differences in the distributions of these two measures. PPD difference scores form a single mode that is relatively sharp near the top, whereas Raw PPD scores are clearly bimodal. This suggests that imitators were tracking the model speaker's alignment as it shifted relative to the nucleus offset with the result that the relative PPD measure appears more stable.

Figure 4: Density plots for two different measures of peak delay: Proportional peak delay (red), and the PPD Difference Score (green).

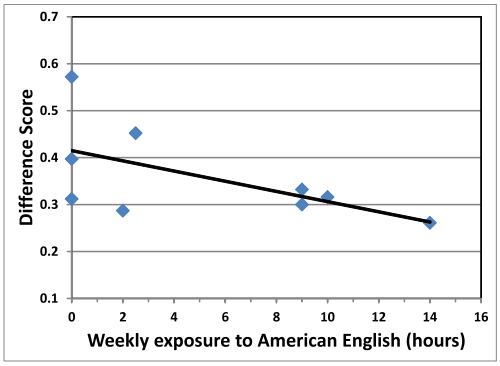


This issue was further addressed by comparing the magnitude (absolute value) of PPD against the difference score across all trials. In a linear mixed effects analysis, there was significant effect of Measure Type (PPD vs. Difference Score) ($\chi^2(1) = 20.99$, $p = 0.001$), with larger magnitudes for PDD values than for Difference Scores. Together, these results show that speakers were tracking the model speaker on a trial-by-trial basis rather than aligning to a fixed segmental landmark such as the nucleus offset.

3.3. Exposure

Subjects' self-reports of weekly exposure to spoken AmE ranged from 0 to 14 hours per week, with a mean of 5.17 hours. The scatterplot in Figure 5 shows how exposure is correlated with the mean of the PPD Difference Score magnitude (absolute value) for each subject. While the present sample is too small to be able to establish either the presence or absence of a relationship, the slight negative correlation between Exposure and Difference Score ($r^2=0.34$) suggests that imitators who have more exposure to AmE patterns are better able to match the *token-by-token* phonetic detail of the AmE productions.

Figure 5: Mean token-specific similarity of PPD to model speaker (Difference Score) by imitator as a function of self-reported weekly exposure to American English.



4. DISCUSSION

The results of the Imitation Task do not seem to support the hypothesis that strong typological differences interfere with Singapore Engish speakers' ability to rapidly adapt and imitate both structural and phonetic detail of the intonational pattern of an AmE speaker. Hence, despite what has been either claimed or implied in some recent cross-dialect imitation studies focusing on intonation and prosodic features [e.g., 5,3,7], though in line with [6], SgE imitators could accurately reproduce phonetic detail relative to the implementation of the tonal contour. As shown above, speakers successfully reproduced L+H* structure and placement, thus suppressing the AP-final rise anchoring that is typical of their variety. Specifically, they were able to shift H peaks earlier in the target region so as to realize them within the stressed vowel nucleus. What is more, proportional alignment values were very close to those of the model AmE speaker. This is somewhat surprising giving that our participants are speakers of a variety of English that is so typologically distinct from the AmE target dialect.

We further showed that speakers were not simply replacing the segmental landmark relative to which H alignment would be implemented (i.e. from word offset to stressed vowel onset), but were actually tracking proportional H alignment within the tone bearing unit as produced by the model speaker on a trial-by-trial basis. Finally, our data suggests that the accuracy in implementing alignment values is correlated with exposure to AmE as measured through a questionnaire. Imtitators with more experience may therefore be better able to attend to the phonetic cues that are relevant to the prosodic system of the target dialect. Future work will establish the relative weight of indexical and native language influence on phonetic imitation accuracy.

8. ACKNOWLEDGEMENTS

This research was funded by a grant from the Singapore Ministry of Education Academic Research Fund Tier 1 (2013-T1-002-169). It was also supported in part by the Erasmus Mundus Action 2 program MULTI of the European Union, grant agreement number 2010-5094-7.

9. REFERENCES

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

[2] Bosshardt, H.-G., Sappok, C., Knipschild, M., and Hölscher, C. (1997). Spontaneous imitation of fundamental frequency and speech rate by non-stutterers and stutterers. *J. Psycholinguist. Res. 26*, 425–448.

[3] Braun, B., Kochanski, G., Grabe, E., and Rosner, B.S. (2006). Evidence for attractors in English Intonation, *JASA, 119(6)*: 4006-4015.

[4] Chong, A. 2013. Towards a model of Singaporean English intonational phonology. *Proc. of the Meetings on Acoustics, 19*.

[5] Cole, J., & Shattuck-Hufnagel, S. (2011). The phonology and phonetics of perceived prosody: What do listeners imitate? In *Proceedings of Interspeech 2011* (pp. 969-972).

[6] D'Imperio, M., Cavone, R. & Petrone, C. (2014). "Phonetic and phonological imitation of intonation in two varieties of Italian". *Frontiers in Psychology*. 2014, 14 pages.

[7] German, J. S. (2012). Dialect adaptation and two dimensions of tune. *Proc. of the 6th International Conference on Speech Prosody* (pp. 430-433).

[8] Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. Psychological Review, 105, 251–279.

[9] Jun, Sun-Ah (2005) Editor. *Prosodic Typology: The Phonology of Intonation and Phrasing.* Oxford University Press.

[10] Nielsen, K.Y. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39 (2)*, 132-142.

[11] Tan, Ying Ying (2010). "Singing the same tune? Prosodic norming in bilingual Singaporeans." in M. Ferreira (ed), *Multilingual Norms*. Frankfurt: Peter Lang.

[12] Wilson, M.D. (1988) The MRC Psycholinguistic

[13] Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers, 20(1)*, 6-11.