Phonetic convergence in temporal organization during shadowed speech

Research Thesis

Presented in partial fulfillment of the requirements for graduation
with *research distinction* in Linguistics in the undergraduate colleges of the Ohio State University

by Roberto Gonzalez

The Ohio State University April 2020

Project Advisor: Dr. Cynthia G. Clopper, Department of Linguistics

Abstract

The goal of this study was to examine phonetic convergence (when one imitates the phonetic characteristics of another talker) in various measures of temporal organization during shadowed speech across different American English dialects. Participants from the Northern and Midland American English dialect regions, plus several "mobile" talkers, were asked to read 72 sentences to establish a baseline for temporal organization, and then to repeat the same 72 sentences after Northern, Midland, and Southern model talkers. Measures of temporal organization (i.e., %V, ΔC, ΔV, rPVI-C, and nPVI-V) were calculated for the read sentences, shadowed sentences, and model talker sentences. Statistical analysis of the differences in distance between the model talker sentences and the shadowers' read and shadowed sentences, respectively, revealed significant convergence by all three shadowing groups toward the model dialects for ΔV, and significant divergence by Mobile talkers away from the model talkers for nPVI-V. Though the result of divergence by Mobile talkers was unexpected, both results provide evidence that support previous studies, which claim that social perception is a large contributing factor in convergence and divergence. These results are also consistent with previous findings demonstrating variation across dialects in temporal organization and, in addition, provide evidence for variation across dialects in convergence in temporal organization.

1. Introduction

Previous research has provided an abundance of evidence that speakers often accommodate their manner of speech to their interlocutors. This accommodative process known as *phonetic convergence* occurs during spontaneous conversations (Gregory Jr. & Webster, 1996; Natale, 1975; Pardo, 2006), and during socially impoverished, laboratory environments such as shadowing tasks (Delvaux & Soquet, 2007; Pardo, 2006; Smith, 2013). The lack of any clear social motivation does not appear to discourage phonetic convergence, and even along with typical shadowing tasks, imitation has also been observed in manipulated speech during shadowing tasks (Brouwer et al., 2010).

It has been argued that convergence is a subconscious, automatic response (Trudgill, 2008), however, divergence has been observed (Babel, 2010, 2012; Giles et al., 1977), showing that there must be factors that either encourage or discourage imitation. Babel (2010, 2012) and Giles et al. (1977) found that social identity is a large contributing factor to imitation or a lack thereof. Babel (2010) summarized the effect of social distance as convergence lessening social distance in between dialects, while divergence maximizes social distance. In both aforementioned studies, social perceptions influenced the magnitude of imitation. When speakers held favorable views of their interlocutors, there was more convergence, but when speakers held negative social views of their interlocutors, either divergence or no significant convergence was observed.

Social selectivity as an effect on imitation is consistent with Clopper & Dossey's (2020) findings. In their study, participants were asked to repeat after or explicitly imitate Southern talkers. While convergence *did* occur in both contexts, there was no significant convergence to the diphthong /aɪ/, as produced by Southern talkers, due to negative perceptions that are held toward this sound in Southern American English. Convergence and divergence do not appear to be automatic responses then, but there are factors that influence how speakers respond to their interlocutors, one of which being social perceptions and identity (Babel, 2009, 2010, 2012; Clopper & Dossey, 2020; Giles et al., 1977; Mitterer & Ernestus, 2008; Mitterer & Müsseler, 2013; Pardo, 2012; Pardo et al., 2017; Walker & Campbell-Kibler, 2015). It's important to note, however, that in socially impoverished environments such as a shadowing task, socially selectivity is still an important factor that effects phonetic convergence, even across difference dialects (Walker & Campbell-Kibler, 2015).

Phonetic distance is another factor that effects convergence. It was observed that imitation is more likely to occur if there is large phonetic distance between the speakers and their interlocutors (Babel, 2010, 2012). While phonetic distance is an important factor that contributes to convergence, observing cross-dialect convergence can be limited by phonetic repertoire (Babel, 2009; Kim et al., 2011). Babel (2009) found that California speaker speakers were more likely to imitate the low vowels of other California speakers as opposed to their high vowels, which is most likely because low vowels allow for more production space (larger phonetic distance allows for larger shifts) while still staying within their phonetic repertoire. Kim et al. (2011) also looked at same-dialect pairs, but also inter-dialect pairs. In that study, they observed that convergence was more likely for same-dialect pairs than for inter-dialect pairs. So, while phonetic distance increases the likelihood of convergence, and convergence has been observed across dialects, too large a distance can inhibit convergence.

Phonetic convergence has been observed concerning f0 (Goldinger, 1997), formants (Babel, 2012; Pardo et al., 2017), intensity (Natale, 1975), consonant articulation (Shockley et al., 2004), vowel duration (Hargreaves, 1960), long-term average spectra (Gregory Jr. & Webster, 1996), and articulation rate (Clopper & Dossey, 2020; Kim et al., 2011). An area to be explored still is temporal organization. Observing convergence in temporal characteristics is particularly interesting because these timing characteristics are unique to each dialect and language (Clopper & Smiljanic, 2015; White et al., 2012) and observing convergence in them sheds more light how people learn, adapt, and accommodate speech. Temporal organization is the distinct, quantifiable timing characteristics of a language (Arvaniti, 2009, 2012; Nolan & Asu, 2009; Wiget et al., 2010). These timing characteristics are the durational patterns of consonants and vowels during speech.

White et al. (2012) found that listeners, when presented with sentences that were "bleached" of specific segments but retained their durational patterns, could differentiate between English and Spanish. While languages have different temporal features that allow speakers to differentiate between them, dialects within American English also have different vocalic and temporal features (Clopper et al., 2005; Clopper & Smiljanic, 2015). Concerning temporal features, Clopper & Smiljanic (2015) found that Midland and Southern speakers have similar speaking rates, while Northerners have faster speaking rates compared to the other two. There are five metrics that are typically used to measure temporal organization and are defined as follows: %V is the proportion of time spent on vowels over the total duration of the sentence, ΔC and ΔV are defined as the standard deviation of the durations of consonant and vowel intervals, respectively, rPVI-C is defined as the average difference between consecutive consonant intervals, and nPVI-V is defined as the average difference between consecutive vowel intervals divided by their average. Concerning the measurements for temporal organization, Southern speakers had the highest proportion of vocalic intervals (measured as %V), while Northern speakers had the smallest. For the standard deviation of consonant intervals (measured as ΔC), Midland speakers had the highest variability, while Southerners had the lowest. For ΔV, Southern and Midland speakers had the greatest variability while Northern speakers had the lowest. rPVI-C showed similar results to ΔC, with Midland speakers having the highest index, while Southern talkers had the lowest. Finally, for nPVI-V, and similarly for ΔV, Midland and Southern talkers had the largest index, while Northern speakers had the lowest. While White et al. (2012) found that people rely on prosodic characteristics to distinguish language, Alcorn et al. (2020) found that in American English, listeners do not rely on prosodic characteristics to distinguish between dialects, but rely on segmental features instead.

Speakers have been shown to accommodate their speech concerning areas like vowel duration and consonant articulation, and convergence has been observed in non-spontaneous speech (e.g. shadowing tasks). Temporal organization is the measurable durational patterns of one's consonant and vowels and there is regional dialect variation concerning temporal features. It is reasonable then to expect that phonetic convergence can be observed in temporal organization during shadowed speech. The purpose of this study was to observe if this was actually the case.

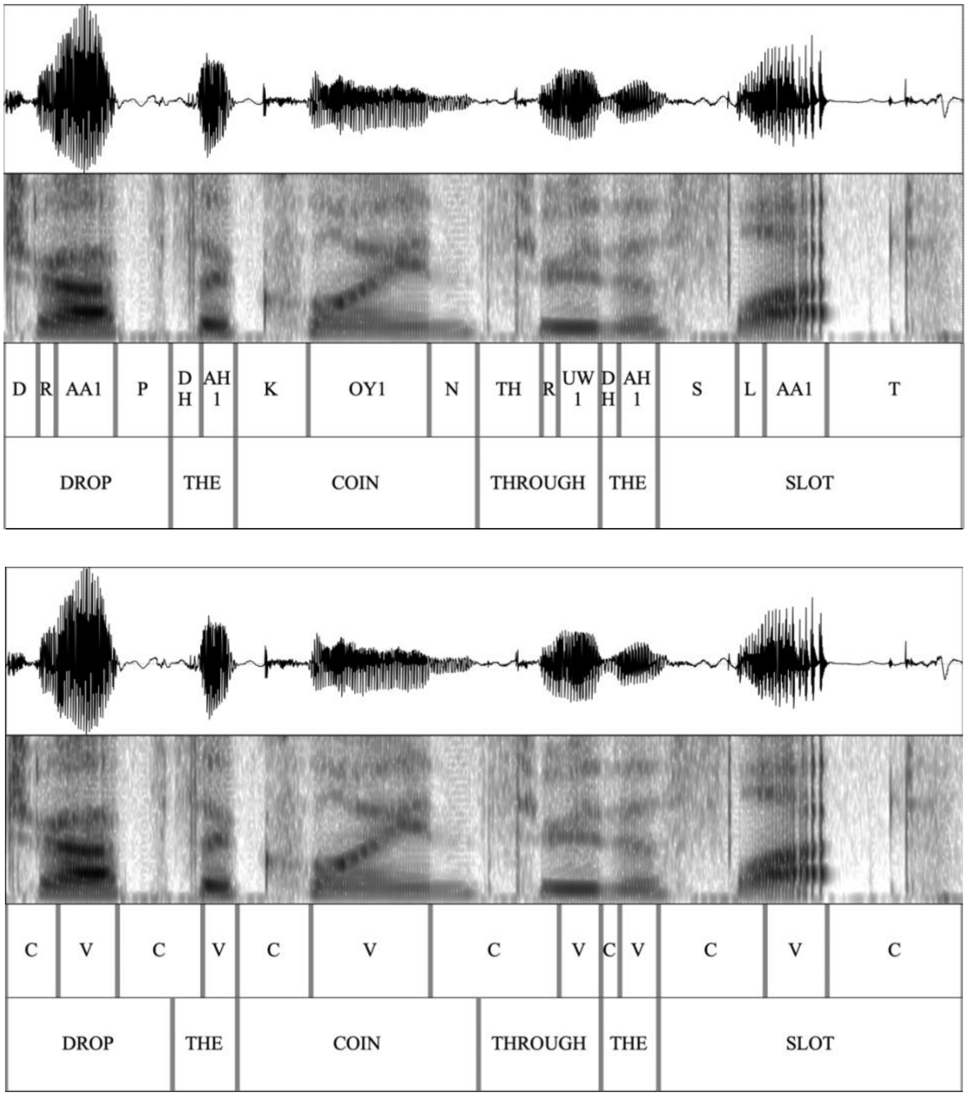
2. Methods

2 .1 Materials

The materials consisted of 72 high predictability sentences produced by two female talkers from each of the Northern, Midland, and Southern dialects of American English (12 sentences each, counterbalanced across six stimulus lists). These six model talkers were native speakers from their respective regions and were speakers from the Indiana Speech Project corpus (Clopper et al., 2002). 10 native female Midland American English talkers, 10 native female Northern American English talkers, and eight female "mobile" talkers (people who have lived in several dialect regions prior to the age of 18) preformed a baseline and then shadowing task using the 72 high predictability sentences. In the baseline task, participants were asked to read the 72 sentences out loud. In the shadowing task, participants were asked to repeat the 72 sentences after the model takers, in which they were not given specific instruction on how to shadow, but simply to repeat. The participants were undergraduate students at the Ohio State University. Only one gender was chosen to participate in this study to reduce variability because different genders have demonstrated different prosodic and temporal features (Clopper & Smiljanic, 2011). Previous research has observed mixed results concerning gender and phonetic convergence. Namy et al. (2002) found that women converge more than men, but Pardo (2006, 2009) found that men converged more than women. In more recent studies, however, Pardo et al. (2017) found that the effect of talker gender was not a significant factor on phonetic convergence, and this conclusion was further supported by Clopper & Dossey (2020).

2.2 Acoustic Measurements

A set of measurements was taken to analyze variation in the duration of vowel and consonant intervals across dialects that included %V, ΔC, ΔV, rPVI-C, and nPVI-V for the model talkers and shadowers during their baseline and shadowed blocks. To obtain these measurements, each file was run through the Penn Forced Aligner wherein all phones were assigned their respective labels, and then the label boundaries were corrected for accuracy. Vowel and consonant intervals were corrected based on the guidelines of Peterson and Lehiste (1960), but unlike their previous guidelines, aspiration following a consonant in the onset position was always considered as part of the consonant, as opposed to sometimes a part the syllable nucleus. All approximants were treated as consonants and differentiated from vowels by a combination of looking at F2 and F3 in the spectrogram and listening to the audio. Upon correction, each phone was relabeled as "V" for vowel, and "C" for consonant. Subsequent vowel and consonant intervals were combined, and their durations extracted for analysis. Figure 1 shows an example of a sentence with individually labeled phones, and then the same sentence with the CV labels where subsequent C and V labels were combined.



%V	ΔC (s)	ΔV (s)	rPVI-C (s)	nPVI-V
0.361	0.049	0.053	0.071	0.766

Figure 1. Example of a sentence that was segmented into its individual phones (top), and then where phones were labeled C for consonant or V for vowel, and subsequent consonant and vowel intervals were combined (middle). Corresponding values for the temporal metrics are shown (bottom).

The complete data set consisted of approximately 67,000 consonant and vowel intervals across 4,464 sentences. 96 sentences were excluded from the shadowers (34 Midland, 38 Northern, and 24 Mobile) due to the participants either misreading the sentences or not repeating after the model talkers correctly.

2.3 Analysis

The goal of this analysis was to determine if shadowers changed their manner of speech from their baseline by converging to the model talkers. The baseline difference for each metric was measured by taking the absolute value of the difference between the model talkers and the baseline readings done by the shadowers for each sentence. Then, the shadowed difference for each metric was measured by taking the absolute value of the difference between the model talkers and the shadowers' repetitions for each sentence. The shadowed difference was subtracted from the baseline difference, and the total difference in distance was found. Positive differences in distance mean convergence to the model talkers due to shadowers starting off with larger differences during their baseline measurements but smaller differences during their shadowed measurements. Negative differences in distance mean divergence from the model talker due to the opposite effect occurring.

One-sample t-tests that compared the values to zero were run to determine if the mean differences in distance were overall different than zero. If the means were zero or near-zero, then the conclusion could be drawn that there was no change from baseline to shadowing, but if the means were significantly different than zero, then the conclusion could be drawn that shadowing had occurred overall. A series of ANOVAs were then run in R to determine if there was a significant effect of shadower dialect on convergence, a significant effect of model dialect on convergence, or if there was significant interaction between the shadower and model dialects. If there were any significant effects or interactions in the metrics, paired t-tests were run to determine where statistically significant differences occurred, in which direction they occurred, and whether that direction was convergence or divergence.

3. Results

3.1 Temporal features for model talkers

Dialect	*%V*	*ΔC* (s)	*ΔV* (s)	*rPVI-C* (s)	*nPVI-V*
Midland	0.398	0.074	0.058	0.080	0.711
Northern	0.405	0.074	0.060	0.078	0.723
Southern	0.408	0.073	0.061	0.081	0.739

Table 1. The mean values of all five temporal metrics for each model talker dialect.

The metrics for the shadower baseline and shadowed blocks were compared to the metrics for the model talkers to identify if convergence to the model talkers occurred. A summary of the metrics for model talker values is in Table 1. The differences in the values in Table 1 are negligible and only differ by a few milliseconds for each metric across the three dialects. Any differences that are pointed out in what follows are not remarkable. What is interesting about these differences, however, is that they *are not* remarkable. One would expect there to be greater variation among these three dialects in each metric. Concerning %V, Midland talkers exhibited the smallest proportion of time spent on vowels, while Southern talkers had the largest proportion of time spent on vowels. These results are different than those of Clopper & Smiljanic (2015), who found that Northern talkers had the smallest proportion of time spent on vowels, but here their %V is higher than the Midland value.

For ΔC, Midland and Northern talkers showed the greatest variability in consonant interval duration, and Southern talkers had the least variability. Clopper & Smiljanic (2015) observed that the relationship between ΔC and %V was negative, meaning that consonant intervals were less variable when vowel intervals were relatively longer. This remains true in these results as well.

ΔV followed the same pattern as %V. While the Southern talkers exhibited the greatest variability in vowel interval duration, Midland talkers had the lowest variability. Northern and Southern talkers only differed in their ΔV values by 1ms, which is a negligible difference. This again differs from Clopper & Smiljanic's (2015) findings where, while Southern talkers had the highest ΔV, Midland talkers also had a relatively high ΔV and Northern talkers had the lowest. Yet here, Midland talkers had a lower variability in vowel interval duration than Northern talkers.

For rPVI-C, Southern talkers had the highest pairwise variability index for consonants, and Northern talkers had the lowest. These results are unexpected when compared to Clopper & Smiljanic (2015), where it was observed that Midland talkers had the largest pairwise consonant interval variability and Southern talkers had the smallest.

For nPVI-V, Southern talkers had the highest pairwise variability index for vowels, and Midland talkers had the lowest. Again, these results are unexpected when compared to Clopper & Smiljanic (2015), though not as much as rPVI-C, where it was observed that Northern talkers had the smallest pairwise vowel interval variability, not Midland talkers. Though these results are not expected, the difference in the means for Midland talkers and Northern talkers was only about 0.01, and the difference in the means for Northern talkers and Southern talkers about 0.01 as well. There was not great variability between the model dialects concerning nPVI-V.

One plausible explanation for why these results are so different than what Clopper & Smiljanic (2015) found could be due to how the data was collected. In Clopper & Smiljanic's (2015) study, speakers read paragraphs as opposed to sentences. The larger stimulus material likely produced different results because as the speakers had more time to adjust to their normal manner of speaking. Reading sentences could have produce homogenous results because there is less room for variation in such a short amount of speaking time.

3.2 Shadower baseline & comparison to model talkers

Dialect	*%V*	*ΔC* (s)	*ΔV* (s)	*rPVI-C* (s)	*nPVI-V*
Midland	0.393	0.080	0.056	0.082	0.652
Mobile	0.582	0.053	0.072	0.075	0.644
Northern	0.552	0.059	0.074	0.075	0.596

Table 2. The mean values of all five temporal metrics for each shadower talker dialect at baseline.

The model talker values were subtracted from the shadowers' baseline values, and the absolute values were taken to establish the baseline difference between the shadowers and the model talkers. All values for the shadowers can be found in Table 2. For %V, the Mobile talkers showed the greatest proportion of time spent on vowels, while Midland talkers exhibited the least. This is an unusual result considering Clopper & Smiljanic's (2015) findings, where it was observed that Northern talkers had the lowest %V, yet here, the proportion of time spent on vowels for Midland talkers was lower by nearly 0.15. The difference between Midland model talkers and Midland shadowers is very small, but Mobile and Northern talkers have much higher %V values than all three model dialects. Because phonetic distance increases the likelihood of phonetic convergence, these differences suggest that Mobile and Northern shadowers are likely to converge to the model dialects. Because all three model dialects have very similar %V values, it's likely that there will be no effect of model dialect on convergence.

For ΔC, Midland talkers had the greatest variability in consonant interval duration, while Mobile talkers had the least variability. Though Clopper & Smiljanic (2015) did not look at Mobile talkers, these results are consistent with their findings that Midland talkers exhibited the greatest ΔC values. These results are also consistent with the negative relationship between %V and ΔC, which states that consonant interval durations are less variable when vowel intervals are longer. The Midland shadowers at baseline were not very different from the model talkers, but Mobile and Northern talkers did differ by greater numbers, with Mobile talkers being the most different from the model talkers. Keeping the effect of phonetic distance on phonetic convergence in mind, it's likely that Mobile and Northern talkers will experience convergence, with Mobile talkers being most likely to experience convergence. Because all three model dialects have very similar ΔC values, it's likely that there will be no effect of model dialect on convergence.

For ΔV, Northern talkers had the greatest variability in vowel interval duration, while Midland talkers had the least variability. This is the same pattern shown by the model talkers, and very unusual considering that the results are reversed in Clopper & Smiljanic's (2015) findings, where it was observed that the Midland talkers exhibited the greatest variability and Northern talkers exhibited the least variability. Compared to the model talkers, Midland shadowers were not that different than all three model talker dialects, but Mobile and Northern talkers had similar and much higher variability than the model talkers. These results suggest that it is likely convergence by the Mobile and Northern shadowers will occur toward the model talkers due to the effect that phonetic distance has on phonetic convergence. Because all three model dialects have very similar ΔV values, it's likely that there will be no effect of model dialect on convergence.

For rPVI-C, Midland talkers had the largest pairwise consonant interval variability and both Mobile and Northern talkers had the smallest. These results are consistent with Clopper & Smiljanic's (2015) findings. Compared to the model talkers, these results are not very different, with the largest difference being the Mobile shadowers, but only by 4-6ms. Due to the lack of phonetic distance between the shadowers and model talkers for rPVI-C, it is unlikely that phonetic convergence will occur concerning this metric.

Finally, for nPVI-V, Midland shadowers had the largest pairwise vowel interval variability and Northern shadowers had the smallest. There is no Southern dialect for the shadowers in this study, but in Clopper & Smiljanic's (2015) findings, Midland talkers had the largest pairwise vowel interval variability just after the Southern talkers. So, these results are consistent with the previous findings where Midland talkers had a relatively high nPVI-V mean (the highest in this current study), and Northern talkers had the lowest nPVI-V mean. Overall, the shadowers had a smaller pairwise variability index for vowels compared to the model talkers, but Northern shadowers showed the most difference from the model talkers, suggesting that the Northern shadowers are most likely to experience convergence due to phonetic distance. Model talker values are very consistent, so there is unlikely to be an effect of model dialect.

3.3 Statistical Analyses

	Midland		Mobile		Northern	
Midland	*%V*	0.0031	*%V*	-0.0028	*%V*	0.0007
	ΔC (s)	0.0025	*ΔC* (s)	-0.0023	*ΔC* (s)	0.0009
	ΔV (s)	0.0014	*ΔV* (s)	0.0021	*ΔV* (s)	0.001
	rPVI-C (s)	0.002	*rPVI-C* (s)	-0.0022	*rPVI-C* (s)	0.0011
	nPVI-V	0.0078	*nPVI-V*	-0.0005	*nPVI-V*	0.0086
Northern	*%V*	0.0043	*%V*	0.0014	*%V*	0.0021
	ΔC (s)	0.0012	*ΔC* (s)	-0.0038	*ΔC* (s)	-0.0022
	ΔV (s)	0.0009	*ΔV* (s)	0.0007	*ΔV* (s)	0.0007
	rPVI-C (s)	0.0018	*rPVI-C* (s)	-0.0062	*rPVI-C* (s)	-0.0006
	nPVI-V	0.0135	*nPVI-V*	-0.0096	*nPVI-V*	0.0075
Southern	*%V*	<0.0001	*%V*	0.0011	*%V*	0.006
	ΔC (s)	0.003	*ΔC* (s)	-0.0033	*ΔC* (s)	0.003
	ΔV (s)	0.0001	*ΔV* (s)	0.0024	*ΔV* (s)	0.0018
	rPVI-C (s)	0.002	*rPVI-C* (s)	-0.0034	*rPVI-C* (s)	0.0034
	nPVI-V	0.0048	*nPVI-V*	-0.0115	*nPVI-V*	0.0151

Table 3. DiD means for each measure collapsed across all model and shadower dialects.

The differences in distance were calculated by subtracting the shadower difference from the baseline difference. The DiD values collapsed across all model and shadower dialects are shown in Table 3. Overall, these differences in distance are very small, which shows that there was not a lot of change by the shadowers. The mix of positive and negative values show that that the changes that did occur were both toward and away from the model talkers, though ΔV was the only metric that saw consistent convergence toward the model takers. Statistical analyses were done to determine if there were significant changes in the shadowers' speech toward the model talkers. First, one-sample t-tests were run to determine if overall convergence or divergence occurred. These analyses showed that there was no significant convergence or divergence for %V, ΔC, rPVI-C, and nPVI-V; however there was significant convergence for ΔV (t(72)=2.9, p= 0.007). So, there was no significant overall change in the shadowers from baseline to shadowing other than for ΔV.

Following the one-sample t-tests, ANOVAs were done on the dependent variable (DiD) with respect to the independent variables (the model dialects and shadower dialects) to determine if there were significant effects of shadower or model dialect, or if there was significant interaction between the shadower dialects and model dialects. ANOVA results show no significant effects or interactions for %V, ΔC, ΔV, and rPVI-C, but there was a marginal effect of shadower dialect for nPVI-V (F(2, 25) = 3, p= 0.066). Two sample t-tests were run to compare the difference in distance means for each shadower dialect to determine which dialects changed more than the others during the shadowing tasks. Northern and Midland were not significantly different than each other, but Mobile was significantly different than both Midland (t(16)=2.1, p= 0.0495), and Northern (t(16)=2.7, p= 0.0148). A one-sample t-test on the Mobile difference in distance mean for nPVI-V confirmed that there was significant difference for Mobile shadowers (t(7)=-2.9, p= 0.0239), but interestingly, the mean was negative (-0.008), which means divergence.

4. Discussion

The results of the analysis show that convergence and divergence occurred in the measures of ΔV and nPVI-V, respectively. Though there were no significant changes to the proportions of vowel duration (%V), the variability of vowel duration intervals did significantly change. Specifically, there was overall convergence in ΔV, and there was divergence by the Mobile talkers in nPVI-V. Consonant metrics did not see any significant change.

The significant overall convergence by all three shadower dialects toward the model dialects for ΔV was expected due to the phonetic distance that existed between the shadowers' baseline and the model talkers. The Northern shadowers saw the biggest change toward the model talkers, which is not surprising because their baseline vowel variability was higher than Clopper & Smiljanic's (2015) findings, but they converged to become more in line with the Northern model talkers, and the model talkers in general. An explanatory variable for why the Northern shadowers had unusually high ΔV values could be that their reading prosody was slower than their regular speaking prosody. It is not surprising, then, that they converged to more typical variability for vowel duration when repeating.

Mobile shadowers experiencing significant divergence in nPVI-V was very unexpected. Mobile shadowers already started with a lower pairwise variability index for vowels than all three model dialects, and during the shadowing tasks they became even lower, diverging from the model dialects. On the surface, this is unusual because there was large phonetic distance between the Mobile shadowers at baseline and model talkers, so convergence should be expected, yet there is a possible explanation for this divergence. The model taker nPVI-V means were relatively high; higher than all the means for nPVI-V found in Clopper & Smiljanic's (2015) findings. It's possible that the Mobile shadowers negatively perceived these large vowel interval variations and reduced their own vowel interval variations in response.

The existence of divergence in this study and the possible explanations for it are supported by Thomas' (2001) findings vowel interval variability contributes to the perception of the Southern drawl, and Clopper & Dossey's (2020) findings concerning how shadowers responded to Southern diphthongs. They found that though speakers accommodated Southern speech, they were unwilling to accommodate the Southern diphthong /aɪ/. With these two ideas in mind, it is possible that there is an overall negative social perception to long vowel interval durations and relatively long diphthongs, and the Mobile talkers responded by diverging from them. The results for Midland and Northern talkers also support this idea because though they did not diverge like the Mobile talkers, they also did not converge. Furthermore, concerning ΔV, it is reasonable to conclude the reason convergence was observed overall is because the vowel variability for the model dialects was lower than the baseline for the shadowers. If lower vowel variability is preferable, then the resulting convergence for ΔV is very reasonable because the shadowers started off with higher ΔV means but converged to the lower ΔV means of the model talkers. The preference for lower vowel variability also explains the divergence by Mobile shadowers away from the model talkers and lack of convergence by Midland and Northern shadowers toward the model talkers for nPVI-V. The model talkers had high nPVI-V means, which are unpreferable. It should also be noted that another reason there is a lack of convergence to highly variable vowel durations is because it could be challenging to do so. It's possible that more consistent vowel interval durations are easier to converge to, but when there is a lot of variation, shadowers aren't able to accommodate to the variation as easily because the changes are inconsistent.

Another explanation for why the imitation patterns were different for ΔV and nPVI-V is that both metrics are trying to explain something slightly different. Though both show variation in vowel interval duration, ΔV looks at variation in vowel interval duration across the entire sentence, while nPVI-V looks at variation in vowel interval duration across consecutive vowel intervals. If the shadowers were disfluent at different points while reading, then this would explain why vowel intervals were more variable overall because speakers stretched out certain vowel intervals and shortened others while reading. Disfluent reading also explains the lower nPVI-V baseline values for shadowers. Consonant onsets are likely places for pauses during reading where the shadowers could have taken a moment to catch up and process the next part of the sentence. Vowel interval durations would be different in the beginning of the sentence compared to the end of the sentence as talkers lengthened syllable nuclei in some places, paused at the onset of the next syllable and then sped up their reading for the blocks of speech in others. So, consecutive vowel intervals would be less variable, but vowel intervals across the entire sentence would be more variable. When listening back to the shadowers baseline exercise, it is clear that there were minor instances of disfluent reading (as mentioned previously, major instances where sentences were misread were removed).

In light of previous studies on the effect of social perception on phonetic convergence and divergence (Babel, 2009, 2010, 2012; Clopper & Dossey, 2020; Giles et al., 1977; Mitterer & Ernestus, 2008; Mitterer & Müsseler, 2013; Pardo, 2012; Pardo et al., 2017; Walker & Campbell-Kibler, 2015), it appears that social perception was the main factor that affected convergence and divergence, and not phonetic distance. It could be argued that there was no convergence to higher vowel interval duration variability because it was not in the shadowers' phonetic repertoire, but there was a consistent pattern of favoring lower variability by the shadowers, and lower vowel interval duration variability appears to be socially preferable, as higher variability could be associated with negative stereotypes.

Acknowledgments

This work was supported by the Ohio State University College of Arts and Sciences Undergraduate Research Scholarship. We would like to thank Joselyn Gilbert, Lauren Eskin, Kenney Hensley, Erin Loxley, and Qingyang Yan for initially collecting the data.

References

Alcorn, S., Meemann, K., Clopper, C. G., & Smiljanic, R. (2020). Acoustic cues and linguistic experience as factors in regional dialect classification. *The Journal of the Acoustical Society of America*, *147*(1), 657–670. https://doi.org/10.1121/10.0000551

Arvaniti, A. (2009). Rhythm, Timing and the Timing of Rhythm. *Phonetica*, *66*(1–2), 46–63. https://doi.org/10.1159/000208930

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351–373. https://doi.org/10.1016/j.wocn.2012.02.003

Babel, M. (2009). *Phonetic and Social Selectivity in Speech Accommodation*. https://escholarship.org/uc/item/1mb4n1mv

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437–456. https://doi.org/10.1017/S0047404510000400

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189. https://doi.org/10.1016/j.wocn.2011.09.001

Brouwer, S., Mitterer, H., & Huettig, F. (2010). Shadowing reduced speech and alignment. *The Journal of the Acoustical Society of America*, *128*(1), EL32–EL37. https://doi.org/10.1121/1.3448022

Clopper, C. G., Carter, A. K., Dillon, C. M., Hernandez, L. R., Pisoni, D. B., Clarke, C. M., Harnsberger, J. D., & Herman, R. (n.d.). *The Indiana Speech Project: An Overview of the Development of a Multi-Talker Multi-Dialect Speech Corpus*. 15.

Clopper, C. G., & Dossey, E. (2020). Phonetic convergence to Southern American English: Acoustics and perception. *The Journal of the Acoustical Society of America*, *ESUSA2020*(1), 671–683. https://doi.org/10.1121/10.0000555@jas.2020.ESUSA2020.issue-1

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*, *118*(3), 1661–1676. https://doi.org/10.1121/1.2000774

Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics*, *39*(2), 237–245. https://doi.org/10.1016/j.wocn.2011.02.006

Clopper, C. G., & Smiljanic, R. (2015). Regional variation in temporal organization in American English. *Journal of Phonetics*, *49*, 1–15. https://doi.org/10.1016/j.wocn.2014.10.002

Delvaux, V., & Soquet, A. (2007). The Influence of Ambient Speech on Adult Speech Productions through Unintentional Imitation. *Phonetica*, *64*(2–3), 145–173. https://doi.org/10.1159/000107914

Giles, H., Taylor, D. M., & Bourhis, R. Y. (1977). Dimensions of welsh identity. *European Journal of Social Psychology*, *7*(2), 165–174. https://doi.org/10.1002/ejsp.2420070205

Gregory Jr., S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, *70*(6), 1231–1240. https://doi.org/10.1037/0022-3514.70.6.1231

Hargreaves, Wm. A. (1960). A Model for Speech Unit Duration. *Language and Speech*, *3*(3), 164–173. https://doi.org/10.1177/002383096000300305

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125–156. https://doi.org/10.1515/labphon.2011.004

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168–173. https://doi.org/10.1016/j.cognition.2008.08.002

Mitterer, H., & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception–action coupling in speech. *Attention, Perception, & Psychophysics*, *75*(3), 557–575. https://doi.org/10.3758/s13414-012-0407-8

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender Differences in Vocal Accommodation: The Role of Perception. *Journal of Language and Social Psychology*, *21*(4), 422–432. https://doi.org/10.1177/026192702237958

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*(5), 790–804. https://doi.org/10.1037/0022-3514.32.5.790

Nolan, F., & Asu, E. L. (2009). The Pairwise Variability Index and Coexisting Rhythms in Language. *Phonetica*, *66*(1–2), 64–77. https://doi.org/10.1159/000208931

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pardo, J. S. (2009). *Expressing Oneself / Expressing One's Self: Communication, Cognition, Language, and Identity*. Psychology Press.

Pardo, J. S. (2012). Reflections on Phonetic Convergence: Speech Perception does not Mirror Speech Production. *Language and Linguistics Compass*, *6*(12), 753–767. https://doi.org/10.1002/lnc3.367

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, *79*(2), 637–659. https://doi.org/10.3758/s13414-016-1226-0

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*(3), 422–429. https://doi.org/10.3758/BF03194890

Smith, B. J. (2013). *The Interaction of Speech Perception and Production in Laboratory Sound Change* [The Ohio State University]. https://etd.ohiolink.edu/pg_10?0::NO:10:P10_ACCESSION_NUM:osu1374116504

Trudgill, P. (2008). Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society*, *37*(2), 241–254. https://doi.org/10.1017/S0047404508080287

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00546

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665–679. https://doi.org/10.1016/j.jml.2011.12.010

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America*, *127*(3), 1559–1569. https://doi.org/10.1121/1.3293004