Comparison of Phonetic Convergence in Multiple Measures

Chelsea Sanker*

1 Introduction

During interaction, speakers acquire characteristics more similar to characteristics of their interlocutors' speech and other behaviors; this is known as *convergence*. In addition to non-linguistic characteristics such as posture (Dijksterhuis and Bargh 2001) and fidgeting movements (Chartrand and Bargh 1999), this has been found in many characteristics of speech, including pitch (Babel and Bulatov 2011), vowel formants (Babel 2012), intensity (Gregory and Hoyt 1982), lexical items (Ireland et al. 2011), syntactic constructions (Nilsenová and Noltig 2010), and timing of conversational turns and pauses (Street 1984).

A variety of explanations for convergence have been proposed. The main explanations for linguistic convergence are understanding-based (e.g. Street and Giles 1982), socially motivated (e.g. Eckert 2001), or automatic (Dijksterhuis and Bargh 2001). Each of these explanations has elements of support from a range of experiments. As more studies add details to the range of influences on convergence and the characteristics affected by it, we can make a clearer picture of the cognitive representations of linguistic characteristics and their dynamics.

The amount of phonetic convergence between speakers has been associated with several factors, such as race (Babel 2012), gender (Pardo 2010), age (Labov 2006), nationality (Giles, Coupland, and Coupland 1991), native language (Kim et al. 2011), and interlocutor status (Gregory and Webster 1996, Bane et al. 2010/2014). One large factor correlated with convergence is positiveness of interlocutors' opinions of each other, using a range of positiveness measures. Convergence occurs to a greater degree between people with positive relationships (Bernieri and Rosenthal 1991) and greater convergence leads to more positive opinions of interlocutors (Giles et al. 1973). Convergence also depends on characteristics of the individuals involved: people who are more concerned with social status exhibit more convergence (Pardo 2006, Natale 1975).

However, the extent to which convergence in different characteristics aligns remains unclear; correlation between convergence in different measures has not been part of many previous studies, in part due to the different experimental designs suited to measuring different characteristics. Convergence in word-level or phoneme-level characteristics such as vowel formants (e.g. Babel 2012) and voice onset time (e.g. Nielsen 2011) or in overall perceived similarity (e.g. Goldinger 1998) have often been featured in shadowing tasks, in which the participants were recorded repeating words immediately after hearing a recording of them (e.g. Goldinger 1998, Babel 2012) and in interactive tasks designed to elicit repetition of words or concepts (e.g. map task in Pardo 2006; mazes in Garrod and Doherty 1994). Larger-scale patterns such as lexical choice, syntax, intensity, turn durations, and pause durations have often been measured in natural conversational settings (e.g. Gregory and Hoyt 1982, Natale 1975, Bernieri and Rosenthal 1991).

While convergence has been observed in each of these measures and some of them are independently correlated with some of the same social conditions, e.g. ratings of closeness and amount in common (Pardo 2012) and absolute measurements, e.g. number of turns and amount of overlapping speech (Levitan et al. 2012), it is not clear that convergence in each of these characteristics behaves the same way, because of the lack of correlation tests between measures. The results of one recent study by Pardo et al. (2015) on F1, F2, and F0 suggest that the pattern of convergence can differ depending on which measure is used. Degree of convergence based on perceptual testing of similarity has been correlated with the convergence exhibited in characteris-

* I would like to thank Sam Tilsen, Abby Cohn, Michael Weiss, and Alan Nussbaum for their help in developing this paper. I am also grateful to Gül Günaydin for sharing her data with me.

tics such as pitch and vowel duration (e.g. Pardo 2010), though in some studies researchers have not found correlation between the results of the perceptual test and the phonetic characteristic tested (e.g. Babel and Bulatov 2011, Pardo et al. 2012).

In this study, the relative patterns of convergence were investigated for eight different characteristics across pairs of interlocutors: F1, F2, vowel duration, F0, intensity, turn duration, duration of pauses marking the transition to a new speaker, and duration of pauses after which the same speaker resumed. Convergence was calculated by comparing the average value in each measure for each participant over four time periods and then comparing the average difference between partners' averages through those time periods. After calculating convergence in each measure, the correlation between convergence in each possible combination of measures was analyzed, to determine how closely convergence in one measure aligns with convergence in other measures.

I hypothesized that convergence in each measure would be positively correlated with the other measure, if the same mechanism underlies convergence in each of these measures; degree of convergence within a pair in vowel formants, F0, intensity, and timing of turns, turn-switching pauses, and within-turn pauses would be predictive of degree of convergence in each of the other measures. The null hypothesis was that convergence in each measure is not correlated with other measures; degree of convergence within a pair in vowel formants, F0, intensity, and timing of turns, turn-switching pauses, and within-turn pauses will not be predictive of degree of correlation in any other measure, suggesting that convergence is mediated through individual differences in attention to different acoustic and timing characteristics of speech, or that there is a slightly different process underlying convergence in different characteristics. Either result has implications for experimental design in convergence research.

2 Methodology

2.1 Participants

There were eight pairs of participants; each pair contained one research assistant (RA) and one volunteer participant; each RA was in two of the pairs, one of high reported liking and one of low reported liking, and each volunteer participant was in a single pair. This produced a total of 12 speakers (4 RAs and 8 volunteer participants), all female speakers of American English, ages 18-22, who reported no speech or hearing difficulties, recruited from the Cornell University campus. These pairs were selected from a larger set of pairs collected for a different study, in order to examine a balanced sample of pairs reporting high and low liking with each RA.

2.2 Procedure

In the first task, the pair was given a sheet of paper with 20 trivia questions taken from the 300 general information questions developed in Nelson and Narens 1980 and told to discuss the questions and write down their best guess for each one. In the second task, the pair was instructed to talk and get to know each other, with no specific directions for discussion topic. An audio and video recording was made of each interaction.

Participants completed a questionnaire after the study, rating their feelings for their partners in a range of aspects. These ratings were averaged to produce a total rating for each individual. In all pairs included here, both participants had given similar ratings of each other. Among these, the highest and lowest scored female-female pair for each RA was used.

2.3 Data Analysis

Recordings were manually labeled in Praat; measurements were taken for duration of labeled sections, pitch and intensity within turns, and formants of vowels. Vowel boundaries were identified based on the audio-video recordings and spectrograms by a labeler blind to the liking-ratings each interlocutor had given her partner. The vowels investigated were /i/, /æ/, /u/, /ɑ/, /ɪ/, and /ʌ/. Vowel formants were identified using Fourier analysis, with Gaussian windows at the midpoint of the vowel, with a length of 0.025 seconds. 1.6% of vowels were omitted because they lacked measurable formants due to rapid speech, low intensity, or co-occurrence with other sounds.

Labels were also given for timing of turn-taking: turns were defined as when a single participant was speaking, pauses with the same speaker before and after the pause were marked as within-turn pauses, and pauses with a different speaker preceding and following them were labeled as turn-switching pauses, and in calculations were associated with the partner who ended the pause.

Before analysis, outliers more than two standard deviations from each speaker's mean for each measure were removed. Vowels from function words were also omitted from calculations, because such vowels are prone to reduction and other effects of context and speech rate (Jurafsky et al. 2001). Furthermore, frequent words can exhibit greater convergence than infrequent words (Goldinger 1998). These factors in combination with the high frequency of function words within the interactions could potentially allow them to skew the data.

3 Results

Despite wide variation in pronunciations, the data reflect significant convergence on average: partners' mean values for measures were becoming more similar, though not all pairs exhibited convergence in all measures. There was a large range in the degree of convergence exhibited in each pair across measures as well as in each measure across pairs, which provided a strong foundation for testing correlations in convergence and change in individuals' means between measures, between tasks, and between pairs with the same RA.

3.1 Convergence within Pairs

The values in each measure were z-normalized in to allow pooling between measures. Convergent change was not present in all measures for all pairs, so the data pooled across pairs does not reflect convergence in some individual measures, although some pairs may have exhibited convergence in those measures; however, convergence was evident in the pooled average among all pairs and all measures. Table 1 gives the average difference, in z-normalized units, between partners' values for each measure at each time quadrant, as well as the pooled average difference.

For the purposes of calculations, the interaction was divided into five-minute quarters, so each quarter reflects half of one of the two tasks; averages were calculated within each quarter. The average difference pooled across all measures and all pairs was significantly reduced from the first to second quarter; using a two-tailed t-test, $t(12424) = 2.31$, $p = .021$. The change from the second to third quarter was also significant: $t(13024) = 2.78$, $p = .0054$. The average change from the third to fourth quarter was not significant: $t(13890) = 0.43$, $p = .66$).

	F1	F2	Vowel Duration	Intensity	F0	Turn Duration	Turn-Switching Pause Duration	Within-Turn Pause Duration	Overall
Q1	0.66	0.28	0.22	0.64	0.31	0.37	0.62	0.42	0.44
Q2	0.56	0.45	0.27	0.48	0.26	0.33	0.42	0.28	0.38
Q3	0.56	0.29	0.19	0.54	0.36	0.28	0.07	0.19	0.31
Q4	0.38	0.37	0.16	0.44	0.39	0.34	0.11	0.20	0.30

Table 1: Average difference between partners for each measure and time quadrant, z-normalized

Vowel measures were z-scored to pool across vowels because most of the individual vowels did not have enough tokens in all pairs to exhibit significant patterns. While none of the differences in convergence for individual vowels were significant, there was a trend towards greater F1 convergence in low vowels, as has been found previously (e.g. Babel 2009), as well as greater F2 convergence in /æ/ and /u/, and greatest convergence in duration for /æ/, and /ɪ/.

There was a significant difference in the average change within the high liking and low liking pairs between all quarters, with a larger decrease in difference between partners in the high liking group. Much of this difference was due to liking-correlated differences in F0, intensity, and turn duration, as seen in Tables 6, 7, and 8. For other measures, there was neither a significant main effect of liking nor significant interaction between liking and other factors.

	High liking	Low Liking	Df	T-value	p-value
Q2-Q1	-0.11	-0.0084	12424	5.50	< .001***
Q3-Q2	-0.052	-0.089	13024	2.09	.038*
Q4-Q3	-0.036	0.011	13890	2.87	.0041***

Table 2: Change in average difference between partners and p-value for difference between the means of each group, z-normalized and pooled among all measures.

These results are consistent with previous work which has found that liking is positively associated with convergence. The presence of a difference only for some of the measures, as is demonstrated in the following section, raises the question of how liking mediates linguistic representations such that it only influences some characteristics.

3.2 Change by Individuals

There was a significant effect of task and time in some measures, with most or all participants exhibiting the same significant shift. There were also several significant differences associated with different liking status and differences by individual pair, and interactions between some of these factors. Tables 3-10 show the p-values for a two-way ANOVA including task and pair for each characteristic.

In the second task, there was a significant increase in average turn duration, a decrease in duration of both types of pause, and an increase in intensity, F0, and F1. Most of these changes seem to be an effect of task and not a direct effect of time, because there is a sharp division in measures between the different tasks but not between the two quadrants within either task. Increased intensity and pitch are associated with greater arousal and task involvement (Street 1984). Longer turns and shorter pauses in the conversational second task than in the trivia game seemed to reflect the speakers' greater confidence in the topic as well as greater interest. Attention has also been identified as a factor which can increase convergence (McIntosh, Druckman, and Zajonc 1994).

	Df	Sum Sq	Mean Sq	F-value	p-value
Task	1	0.52	0.52	16.8	< .001***
Liking	1	0.0015	0.0015	0.05	.83
Pair	6	0.62	0.1	3.3	.026*
Task+Liking	1	0.062	0.062	2.0	.18
Task+Pair	6	0.35	0.058	1.9	.15
Residuals	16	0.5	0.031		

Table 3: ANOVA for effect of task, liking, and pair: F1

	Df	Sum Sq	Mean Sq	F-value	p-value
Task	1	0.0041	0.0041	0.085	.77
Liking	1	0.085	0.085	1.76	.20
Pair	6	2.07	0.34	7.1	.0080***
Task+Liking	1	0.13	0.13	2.6	.13
Task+Pair	6	0.36	0.59	1.22	.34
Residuals	16	0.78	0.049		

Table 4: ANOVA for effect of task, liking, and pair: F2

	Df	Sum Sq	Mean Sq	F-value	p-value
Task	1	0.3	0.3	0.0031	.96
Liking	1	127.6	127.6	1.33	.27
Pair	6	2120.4	353.4	3.67	.017*
Task+Liking	1	0.032	0.032	0.00033	.99
Task+Pair	6	385.4	64.2	0.67	.67
Residuals	16	1540.1	96.26		

Table 5: ANOVA for effect of task, liking, and pair: Vowel duration

	Df	Sum Sq	Mean Sq	F-value	p-value
Task	1	3239.9	3239.9	8.56	.0099**
Liking	1	52682.5	52682.5	139.2	< .001***
Pair	6	284191.4	47365.2	125.2	< .001***
Task+Liking	1	81.4	81.4	0.22	.65
Task+Pair	6	4297.7	716.3	1.89	.14
Residuals	16	6053.5	378.3		

Table 6: ANOVA for effect of task, liking, and pair: F0

	Df	Sum Sq	Mean Sq	F-value	p-value
Task	1	922.9	922.9	18.3	.00058***
Liking	1	63979.1	63979.1	1269.2	< .001***
Pair	6	383036.7	63839.5	1266.5	< .001***
Task+Liking	1	560.3	560.3	11.1	.0042**
Task+Pair	6	3770.7	628.4	12.5	< .001***
Residuals	16	806.5	50.4		

Table 7: ANOVA for effect of task, liking, and pair: Intensity

	Df	Sum Sq	Mean Sq	F-value	p-value
Task	1	2565090.1	2565090.1	46.6	< .001***
Liking	1	314384.3	314384.3	5.7	.029*
Pair	6	2237168.3	372861.4	6.78	.001**
Task+Liking	1	38707	38707	0.7	.41
Task+Pair	6	1390393.9	231732.3	4.21	.0099**
Residuals	16	880340.3	55021.3		

Table 8: ANOVA for effect of task, liking, and pair: Turn duration

	Df	Sum Sq	Mean Sq	F-value	p-value
Task	1	4052648.6	4052648.6	19.3	.00045***
Liking	1	51096.5	51096.5	0.24	.63
Pair	6	555370	92561.7	0.44	.84
Task+Liking	1	12483.6	12483.6	0.06	.81
Task+Pair	6	145453.7	24242.3	0.12	.99
Residuals	16	3355813	209738.3		

Table 9: ANOVA for effect of task, liking, and pair: Turn-switching-pause duration

	Df	Sum Sq	Mean Sq	F-value	p-value
Task	1	1593514.7	1593514.7	10.9	.0044*
Liking	1	27218.2	27218.2	0.19	.67
Pair	6	692568.5	115428.1	0.79	.59
Task+Liking	1	112.1	112.1	0.00077	.98
Task+Pair	6	251026	41837.7	0.29	.93
Residuals	16	2329916.5	145619.8		

Table 10: ANOVA for effect of task, liking, and pair: Within-turn-pause duration

The different effect of task and liking in different measures indicates the importance of considering how observed patterns of convergence might be reflecting influences of the task design; given how differently the measures pattern in the two tasks, looking at convergence in a single task may present a misleading picture of the behavior of some characteristics. Because of the significant effect of time on the means of some measures, many of the subsequent tests of correlation significance will compare change between individuals who were not interacting, in order to provide a baseline for correlations due to these shifts.

Despite shifts in many characteristics, there was consistency in individuals' pronunciations: for speakers' mean values across all measures in the first vs. second task, r(126) = 0.53, $p < .001$. For the two halves of the first task, r(126) = 0.67, $p < .001$; for the two halves of the second task, r(126) = 0.74, $p < .001$. That is, speakers' means for each measure in each time quadrant were correlated with their means for the same measure in the other quadrants, although they exhibited some changes due to task and time. Most measures underwent significant change for many individuals; the mean change over time across all measures and individuals was 0.41 (SD = 0.32), significantly higher than 0 (i.e. if pronunciations were not changing), t(127) = 14.5, $p < .001$.

	F1	F2	Vowel Duration	Intensity	F0	Turn Duration	Turn-Switching Pause Duration	Within-Turn Pause Duration
Q2-Q1	0.18	0.21	0.3	0.098	0.32	0.29	0.23	0.18
Q3-Q2	0.26	0.20	0.28	0.27	0.80	0.51	0.29	0.62
Q4-Q3	0.19	0.25	0.18	0.17	0.14	0.13	0.20	0.15

Table 11: Average change by individuals, z-normalized

However, convergence is not the result of a consistent shift of each speaker's means in the direction of her partner's mean for each measure. The correlation between the change in difference between partners and the change of the individuals within a pair does not reach significance. For all measures and pairs, r(190) = -0.12, $p = .095$; for only measures not associated with turn-taking and thus less directly influenced by change in task, r(118) = 0.15, $p = .098$.

	F1	F2	Vowel Duration	Intensity	F0	Turn Duration	Turn-Switching Pause Duration	Within-Turn Pause Duration
r(24)	0.0056	0.34	0.21	0.10	0.31	-0.16	-0.53	-0.32
p-value	0.98	0.085	0.31	0.64	0.12	0.45	0.0022**	0.11

Table 12: Correlation between change made by the individuals of a pair across each time period and change in the difference between those partners; p-values for the correlations.

The significant negative correlation which appears for turn-switching pause duration is reflecting the effect of a strong shared shift of all participants towards shorter pauses over time, particularly between the first and second task. This time-related individual change had a strong influence on difference between partners' averages for this measure; turn-switching pauses were longest at the beginning of the interaction, which also allowed for more variation and greater difference between partners as well as providing room for change, while at the end of the interaction turn-switching pauses were short for all participants, so differences between partners were small and speakers were changing their pause durations less because there was little room for change.

The change made by an individual was strongly correlated with the change made by her partner in the same measure over the same time period: $r(190) = 0.53$, $p < .001$. In part this correlation was due to patterns of certain characteristics over time across participants, which is reflected in a high correlation between mean change in each characteristic and time period between individuals from different pairs in the same liking condition; for these comparisons, the correlation was lower but still significant: $r(190) = 0.29$, $p < .001$. However, the difference between these correlations was statistically significant $Z = 2.83$, $p = .0047$. The higher correlation among partners than among individuals who were not interacting with each other may suggest that interlocutors are responding to the variability of their partner's characteristics, though it may in part also be an indirect consequence of the shared degree of positiveness of the interlocutors' relationships.

For some individual measures, the difference in correlation coefficients also reaches or approaches significance; the comparison for each measure is given in Table 13. Due to strong patterns associated with change across the tasks and change particular to each measure, which results in correlations that are independent of convergence, the correlation in change within pairs is compared to the correlation between non-partners using Fisher's z transformation.

	F1	F2	Vowel Duration	Intensity	F0	Turn Duration	Turn-Switching Pause Duration	Within-Turn Pause Duration
r(22) of partners	0.41	0.27	0.31	0.72	0.34	0.34	0.57	0.52
r(22) of non-partners	-0.16	-0.23	0.086	0.21	-0.071	0.021	0.58	0.26
Z-score for difference	1.93	1.66	0.76	2.25	1.38	1.08	-0.05	1.01
p-value of difference	.054	.097	.45	.024*	.17	.28	.96	.31

Table 13: Correlation between partners' average change in each characteristic; correlation between change by randomly matched non-partners with the same liking level; p-value for the difference between those correlations.

3.3 Factors: Patterns of a Particular Pair or Individual

There were correlations between the tasks based on particular pairs and individuals, which suggests that individuals are somewhat consistent in the degree to which they exhibit change in each characteristic and in the degree of convergence which their pairs exhibit in each measure. The correlation in convergence in each measure within pairs which contained the same individual was significantly higher than the correlation in convergence in each measure between pairs which did not contain the same individual. In addition, the amount of change exhibited by each individual in one task was significantly correlated with the amount of change which that individual exhibited in the second task and significantly greater than correlations between change from different individuals in the two tasks.

The amount of convergence exhibited by a pair in the same measure during different tasks was significantly correlated, depending on which measures were included in calculations. Correlations between measures within each task within pairs were compared to correlations between measures within each task for different pairs, to avoid correlations which were due to patterns of change over time exhibited by all pairs and not characteristics specific to the pairs.

Turn-taking measures seemed to be more strongly affected by task than by the individuals involved. Among all measures, including turn-taking characteristics, r(62) = -0.22, p = .074 within the same pair and r(62) = 0.076, p = .55 for comparing the change in one task from one pair and in the other task from a different pair. The difference between these correlations does not reach significance: Z = -1.66, p = 0.097. The correlation comparisons for individual measures are given in Table 14.

	F1	F2	Vowel Duration	Intensity	F0	Turn Duration	Turn-Switching Pause Duration	Within-Turn Pause Duration
r(6) for same pair	0.72	-0.24	0.66	-0.30	0.0015	-0.78	-0.014	-0.33
r(6) for different pairs	-0.16	-0.52	0.23	0.56	0.18	0.28	-0.020	0.13
Z-score for difference	1.69	0.52	0.88	-1.49	-0.29	-2.11	0.01	-0.75
p-value of difference	.091	.60	.38	.14	.77	.035*	.99	.45

Table 14: Correlation between change in difference between partners' means (convergence) during the first task and during the second task; correlation between convergence within the first task and the second task from different pairs; p value for the difference between those correlations.

The negative correlation for tasks within the same pair is largely due to the strong negative correlation in turn duration. This negative correlation for turn durations between tasks within the same pair may be because pairs which underwent the most timing change over the course of the first task were most likely to have reached a more stable timing pattern by the second task.

Among only the measures which are not associated with turn-taking and thus are less directly influenced by task type, r(38) = 0.49, p < .001 within the same pair and r(38) = 0.097, p = .44 for comparing the change in one task from one pair and in the other task from a different pair. The difference between these correlations is marginally significant: Z = 1.89, p = .059, which suggests that at least when looking just at measures which are largely task-independent, certain pairs are more prone to convergence in particular measures across different tasks than different pairs, particularly when these tendencies are compared across measures.

The observed convergence in different tasks within a pair was significant when compared to tasks from different pairs. Within the same pair, r(126) = 0.15, *p* = .087; with change in each task from a different pair, r(126) = -0.14, *p* = .11. The difference between these correlations is significant: Z = 2.31, *p* = .021. However, none of the individual measures reach significance in the comparison of correlations within a pair and with each task from a different pair, as seen in Table 15.

	F1	F2	Vowel Duration	Intensity	F0	Turn Duration	Turn-Switching Pause Duration	Within-Turn Pause Duration
r(14) for same pair	0.34	-0.057	0.024	0.21	0.35	-0.072	-0.011	0.55
r(14) for different pairs	-0.19	-0.075	-0.23	-0.039	-0.25	0.10	0.39	0.25
Z-score for difference	1.39	0.05	0.66	0.64	1.58	-0.44	-1.08	0.93
p-value of difference	.16	.96	.51	.52	.11	.66	.28	.35

Table 15: Correlation between individual change during the first task and during the second task; correlation between change in the first task and the second task by individuals from different pairs; p value for the difference between those correlations.

The correlation between convergence for pairs including the same individual as compared to pairs not including the same individual was significant, although the only individual measure for which it reached significance was intensity, as seen in Table 16. Because each RA was included in one pair with high liking ratings and one with low liking ratings, any correlation appears despite the competing effect of liking.

	F1	F2	Vowel Duration	Intensity	F0	Turn Duration	Turn-Switching Pause Duration	Within-Turn Pause Duration
r(14), pairs with shared person	0.11	0.30	0.019	0.81	0.41	0.40	0.45	-0.065
r(14), pairs with no overlap	-0.11	-0.19	-0.050	-0.49	0.14	0.71	0.45	-0.080
Z-score for difference	0.56	1.28	0.18	4.24	0.75	-1.18	0	0.04
p-value of difference	.58	.20	.86	<.001***	0.45	.24	1	.97

Table 16: Correlation between convergence between pairs containing one of the same individuals; correlation between convergence between pairs with no shared individual; p value for the difference between those correlations.

For the overall convergence between pairs with the same RA, r(126) = 0.48, *p* < .001; for overall convergence between pairs with no individual in common, R = 0.032, *p* = .72. The difference between these correlation coefficients is significant: Z = 3.88, *p* < .001. This suggests that particular individuals tend to converge on particular characteristics, though it must be considered that convergence is the result of interaction between individuals, so convergence in these pairs cannot just be attributed to change by this shared individual.

The change made by each individual was correlated with the change made by that individual in a different pair of the other liking condition r(94) = 0.31, p = .0013; however, the correlation between change made by different individuals in different liking conditions is r(94) = 0.12, p = .24, and the difference between these correlations is not significant: Z = 1.36, p = 0.16, which makes the significance of the correlation in change made by an individual less interpretable. The comparison in correlation coefficients does reach significance for some measures individually, as seen in Table 17.

	F1	F2	Vowel Duration	Intensity	F0	Turn Dura-tion	Turn-Switching Pause Duration	Within-Turn Pause Duration
r(10) for same person	-0.58	0.24	0.0	0.35	-0.30	0.32	0.71	0.70
r(10) for dif-ferent people	0.30	0.047	0.30	-0.025	-0.66	0.26	0.43	-0.056
Z-score for difference	-2.06	0.42	-0.66	0.83	1.03	0.14	0.91	1.96
p-value of difference	.039*	.67	.51	.41	.30	.89	.36	.050*

Table 17: Correlation between individual change by the same person in different pairs across each time period; correlation between change by different non-interacting individuals; p value for the difference between those correlations.

The overall lack of higher correlations between change made by the same individual in different pairs than between change made by different individuals may be due to the difference in liking in each of the pairs which each RA participated in; this suggests that the degree of change which an individual exhibits in speech characteristics is strongly influenced by how positive a particular interaction is, though there may also be a trend for individuals to be more or less prone to certain types of change in particular characteristics, as suggested by the correlation in convergence between pairs with the same individual.

3.4 Correlations in Convergence

There was little correlation between the degree of convergence in different measures for each pair and there was no trend for correlations to be positive or negative; the amount of change between partners exhibited an only slightly higher trend towards positive correlation.

Among the 28 correlations between convergence in different measures, there were only four correlations which reached significance, all of which can be attributed to changes shared by most speakers across the course of the interaction, independent of partners' characteristics, as the comparisons were made between measures for each time period.

The amount of change by partners within a pair in different measures exhibited a slight trend towards positive correlation, though this comparison also resulted in three significant correlations, at least two of which have likely independent explanations. The pairs of measures which have significant correlations for convergence have no overlap with the pairs of measures which have a significant correlation in individual change.

Correlations between convergence in different measures are given in Table 18. There is no trend for these correlations to be positive (using Fisher's z transformation, the mean correlation r(22) = -0.014, p = .95), and the majority of the individual correlations also do not reach significance. The four pairs of measures for which the correlation did reach significance are: vowel duration and turn-switching pause duration; vowel duration and within-turn pause duration; turn duration and within-turn pause duration; turn-switching pause duration and within-turn pause duration.

	F2	Vowel Duration	Intensity	F0	Turn Duration	Turn-Switching Pause Duration	Within-Turn Pause Duration
F1	-0.086	-0.15	0.42	-0.089	-0.26	0.21	-0.049
F2		0.21	-0.093	0.023	0.015	0.26	-0.0090
Vowel Duration			-0.34	-0.027	0.24	-0.44*	-0.43*
Intensity				-0.036	0.098	-0.078	0.065
F0					-0.060	0.024	-0.22
Turn Duration						0.058	0.42*
Turn-Switching Pause Duration							0.44*

Table 18: Correlations for change in difference in different characteristics across time (Df = 22)

The correlation between convergence in within-turn pause duration and turn switching pause duration is largely based on task-related change, which is evident when comparing correlation in change in difference in both measures within the same pair to the correlation in change in difference with each measure taken from a different pair. Within all pairs, average durations of both types of pause decreased, resulting in more similar pause durations across all speakers. For the mean change in difference in within-turn and turn-switching pause durations within the same pair, r(22) = 0.44, p = .018; for the mean change in difference in within-turn and turn-switching pause durations in different pairs, r(22) = 0.28, p = .17. The difference between these correlation coefficients is not significant: Z = 0.6, p = 0.55.

The correlation between change in difference between partners' mean turn duration and within-turn pause duration is exhibits the same pattern. For the mean change in difference in turn durations and within-turn pause durations within the same pair, r(22) = 0.42, p = .026; for the mean change in difference in turn durations and within-turn pause durations in different pairs, r(22) = 0.18, p = .39. The difference between these correlation coefficients is not significant: Z = 0.86, p = 0.39

The correlation between convergence in vowel duration and within-turn pause duration can be similarly attributed to change across time, with vowels getting shorter and pauses also getting shorter. For the mean change in difference in vowel durations and within-turn pause durations within the same pair, r(22) = -0.43, p = .022; for the mean change in difference in vowel durations and within-turn pause durations in different pairs, r(22) = -0.46, p = .012. The difference between these correlation coefficients is not significant: Z = 0.12, p = .9.

Likewise the correlation between convergence in vowel duration and turn-switching pause duration can be attributed to change across time, with vowels getting shorter and pauses getting shorter. For the mean change in difference in vowel durations and turn-switching pause durations within the same pair, r(22) = -0.44, p = .018; for the mean change in difference in vowel durations and turn-switching pause durations in different pairs, r(22) = -0.37, p = .057. The difference between these correlation coefficients is not significant: Z = -0.27, p = 0.79.

Change by speaker in each measure was also generally not significantly correlated with change by that speaker in most other measures, as seen in Table 19. There was only a very weak trend towards positive correlation (using Fisher's z transformation, the mean correlation r(46) =

0.11, *p* = .45); many of the correlations were small, but few of them were strongly negative. The correlation reached significance for three pairs of measures: intensity and turn-switching pause duration; intensity and within-turn pause duration; and intensity and F1. The first two of these can be attributed to effects of task or time, as there was no significant difference between the correlation between change in the characteristics by the same individual or by different individuals for each characteristic. The remaining correlation lacks a clear explanation.

	F2	Vowel Duration	Intensity	F0	Turn Duration	Turn-Switching Pause Duration	Within-Turn Pause Duration
F1	0.070	-0.093	0.33*	0.051	0.27	0.20	0.12
F2		-0.086	0.021	-0.13	0.048	0.067	0.075
Vowel Duration			-0.033	-0.070	0.075	0.020	0.14
Intensity				0.030	0.19	0.56*	0.56*
F0					-0.21	0.27	-0.14
Turn Duration						0.25	0.16
Turn-Switching Pause Duration							0.27

Table 19: Correlations between change in different characteristics across time (Df = 46)

Although the correlation between change in mean intensity and turn-switching pause duration for the same individual was significant: r(46) = 0.56, $p < .001$, this result can be attributed to task-related changes, as the same pattern is present in the change exhibited by different individuals for these measures: r(46) = 0.59, $p < .001$. The difference between these correlation coefficients is not significant: Z = -0.21, *p* =.83.

The correlation between change in average intensity and within-turn pause duration also likely attributable to task-related changes. While the correlation between change in intensity and change in within-turn pause duration for the same individual is significant: r(46) = 0.56, $p < .001$, an only slightly weaker pattern is present in the change exhibited by different individuals for these measures: r(46) = 0.27, *p* = .054. The difference between these correlation coefficients does not reach significance: Z = 1.69, *p* =.091.

The correlation between change in average intensity and F1 is significant for the same individual: r(46) = 0.33, *p* = .016, while the correlation between these measures from different individuals is not: r(46) = -0.08, *p* = .59. The difference between these correlation coefficients is significant: Z = 2.01, *p* = .044. The lack of clear external influences responsible for this effect may suggest that it reflects a real connection between variability within these measures. However, the lack of strong correlation in any other pairs of measures makes this correlation much less compelling; such a result could easily arise by chance given the large number of correlations being considered here. Nevertheless, it would be valuable to further investigate the possibility that speakers who have more variable productions in one characteristic may also be somewhat more prone to variation in other characteristics; if such a tendency is present but overshadowed in these results by other influences such as differences based on task, it may reach significance in a larger data set.

4 Discussion

The results of this study show a correlation between the amount of change exhibited by the same individual in the same speech characteristic during different tasks, as well as between the amount of convergence exhibited in different pairs containing the same speaker, but not in the amount of change exhibited by the same individual in the same characteristic in different pairs. Furthermore, there was no clear pattern of correlation between convergence in different measures within a pair or each individual's change in different measures within a pair.

The results of this study, with convergence behaving differently in different measures, have bearing on the issue of how to characterize linguistic memory. Some investigators (e.g. Goldinger 1998) argue for a purely episodic memory, in which details of specific instances of hearing a word or phoneme are stored, with greater salience of recent exemplars. Exemplar models are generally good at explaining convergence, but do not provide a model for why convergence should differ by the characteristic measured. The results of this study are most consistent with a model which includes exemplar details, but also contains a system of abstractions, such that speakers can differ in how they weight exemplars for different characteristics or in whether the surface forms of a characteristic are produced more through the application of regular implementation rules (Cf. Pierrehumbert 2002) or more based on recent exemplars.

The differences in convergence in different characteristics have implications for modeling the process of convergence. Linguistic convergence has sometimes been explained as being an innate process (e.g. Dijksterhuis and Bargh 2001). In support of this, several experiments have shown that convergence cannot be reliably controlled consciously, either to avoid it (Lewandowski 2012) or increase it (Pardo 2010), though people are somewhat able to influence their degree of convergence indirectly, based on attempting to create a particular social impression (Putman and Street 1984). Convergence has also been attributed to social strategy, because people perceive similarity to themselves favorably and are thus likely to behave more similarly to people by whom they want to be favorably perceived. Dijksterhuis and Bargh (2001) attribute differences in imitation based on the positiveness of relationships as an effect of attention, because people pay more attention to speakers whom they like and are more likely to be influenced by input which they have been attending to. Also consistent with this explanation is the finding that people who are more empathetic or concerned with social position display more mimicry (Nguyen, Dufour, and Brunellière 2012). Convergence has alternatively been described as functioning to promote understanding. Imitation seems to improve understanding of a new accent; people perform best at understanding sentences in an unfamiliar dialect partially masked by noise if they had practice imitating it (Adank and Bekkering 2010). However, shift in pronunciation is not always associated with a shift in underlying representations (Evans and Iverson 2007).

The existence of evidence supporting each of these different theories about the process of convergence may in part be due to differences in convergence among different measures. The scant support found by this study for correlation between convergence in different measures suggests that convergence in different characteristics must be mediated by personal experience, attention, and other factors, or even be a different process in different characteristics. Perhaps different motivations driving interactions, including factors such as the positiveness of the interlocutors' relationship and the context of the interaction, can influence which characteristics exhibit the most convergence, if convergence is mediated by a number of factors such as understanding and social goals.

This individual variation may interact with whether variability within a characteristic is part of the mental representation of a sound modulated by the phonological environment (e.g. vowel formants) or whether it is largely controlled by aspects of the social setting and other external factors (e.g. turn duration). Several of the measures in the latter group are associated with emotional arousal and attention (Goudbeek and Scherer 2010; Street 1984), and thus could be expected to pattern together. However, within these results, the patterns of correlations between measures are not significantly different for the non-phonological measures (turn taking, pause durations, F0, and intensity than for phonological measures (F1, F2, vowel duration). There is a slightly higher mean correlation in convergence in non-phonological measures ($r(22) = 0.076$, $p = .72$) than in phonological measures ($r(22) = 0.007$, $p = .97$), but the difference is not significant:

Z = 0.27, $p = 0.79$. There is a higher mean correlation in individual change in non-phonological measures (r(46) = 0.21, $p = .14$) than in phonological measures (r(46) = -0.036, $p = .81$); however, this difference also does not reach significance: Z = 1.18, $p = .24$. The lack of significance makes this aspect of the results resistant to interpretation; if the differences become more apparent in a larger data set, this may suggest that because individual differences in phonological representations are relatively small, convergence in phonological characteristics is more constrained and thus likely to differ less between individuals, while non-contrastive characteristics can be more variable and thus better reflect individuals' proneness to change and convergence.

The results from this experiment demonstrate the need for further experiments to investigate individual differences in convergence in different characteristics, in which each participant performs the experiment multiple times, with different partners. The findings should be replicated or expanded with a larger sample size and with participants who are not research assistants, to confirm that there really are individual differences in which characteristics particular individuals are prone to convergence in.

Furthermore, additional work to establish the patterns of individual differences in variability and convergence in different linguistic characteristics and to develop potential explanations for these patterns would be valuable. Greater convergence in certain measures may be due to greater salience of these characteristics in interlocutors' speech as phonetic or conversational cues (e.g. Hazan and Rosen 1991), with convergence on each measure reflecting the interaction of cues which each partner is attending to. Because different people are more sensitive to different speech characteristics (Yu 2013), these are likely to be the characteristics that they are most likely to attend to in the speech of their interlocutors and subsequently exhibit imitation in.

This could perhaps be investigated in an experiment testing whether the characteristics which an individual exhibits the most convergence in are also the characteristics whose absence is most detrimental to that individual identifying a sound or speaker. In addition, there may be a correlation between predisposition to convergence or change in certain characteristics associated with particular personality traits, which may be investigated by including a test of e.g. empathy (Cf. Chartrand and Bargh 1999) or social desirability (Cf. Pardo 2006, Natale 1975) in conjunction with a study on patterns of convergence exhibited by individuals.

5 Conclusions

The results demonstrate a lack of correlation between convergence in different phonetic measures, aside from correlations due to physiological relationships and task-related changes. Despite convergence to varying degrees exhibited in the characteristics measured, convergence of a pair of interlocutors in one measure was not correlated with convergence in other measures.

The correlations between convergence by the same pair during different tasks and by pairs containing the same individual, which were higher than the correlations between different pairs or pairs with no shared individual, suggest that the absence of correlations between measures is not the result of limitations in the study in the small number of pairs, the brief interaction, or the variability inherent in the conversational task. Based on these results, I am inclined to provisionally accept the null hypothesis, that convergence in each measure is not correlated with convergence in other measures. This warrants further investigation of convergence in multiple measures, to clarify the effects of task and particular speaker on convergence in different characteristics.

Speakers' change in a given measure in one task was significantly correlated with their change in that measure in a second task, to a greater degree than the correlation between change made by different speakers in the two tasks. However, change of a speaker's productions of these characteristics was not strongly associated with change in other characteristics, beyond correlations due to shifts exhibited by all speakers and a slight trend for positive correlations; these results neither strongly support the hypothesis of individual tendency towards more or less stable speech characteristics nor clearly disprove it. Different individuals may be more prone to variation in different characteristics, though it remains unclear just how this variability is associated with convergence in these characteristics.

The lack of correlation found in this study between convergence in different measures indicates that patterns of convergence in one measure cannot be taken to be representative of overall convergence, which opens up an important aspect to be considered in future studies on convergence.

References

Adank, Patti, Peter Hagoort, and Harold Bekkering. 2010. Imitation improves language comprehension. *Psychological Science* 21: 1903–1909.

Babel, Molly. 2009. Phonetic and Social Selectivity in Speech Accommodation. Doctoral dissertation, University of California at Berkeley.

Babel, Molly. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40: 177–189.

Babel, Molly, and Dasha Bulatov. 2011. The role of fundamental frequency in phonetic accommodation. *Language and Speech* 55.2: 231–248.

Bane, Max, Peter Graff, and Morgan Sonderegger. 2014. Longitudinal phonetic variation in a closed system. In *Proceedings of the 46th Annual Meeting of the Chicago Linguistics Society*, ed. R. Baglini, T. Grinsell, J. Keane, A.R. Singerman, and J. Thomas, 43–58.

Bernieri, Frank, and Robert Rosenthal. 1991. Interpersonal coordination: Behavioral matching and interactional synchrony. In *Fundamentals of Nonverbal Behavior*, ed. R.S. Feldman and B. Rime, 401–432. Cambridge: Cambridge University Press.

Chartrand, Tanya, and John Bargh. 1999. The chameleon effect: The perception behavior link and social interaction. *Journal of Personality and Social Psychology* 76: 893–910.

Dijksterhuis, Ap, and John Bargh. 2001. The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology* 33: 1–40.

Eckert, Penelope. 2001. Style and social meaning. In *Style and Sociolinguistic Variation*, ed. P. Eckert and J. Rickford, 85–108. Cambridge: Cambridge University Press.

Evans, Bronwen, and Paul Iverson. 2007. Plasticity in Vowel perception and Production: A Study of Accent Change in Young Adults. *Journal of the Acoustical Society of America* 121.6: 3814–3826.

Garrod, Simon, and Gwyneth Doherty. 1994. Conversation, co-ordination, and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition* 53: 181–215.

Giles, Howard, Justine Coupland, and Nikolas Coupland. 1991. *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press.

Giles, Howard, Donald Taylor, and Richard Bourhis. 1973. Towards a theory of interpersonal accommodation through language: some Canadian data. *Language in Society* 2: 177–192.

Goldinger, Stephen. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105: 251–279.

Goudbeek, Martijn, and Klaus Scherer. 2010. Beyond arousal: Valence and potency/control in the vocal expression of emotion. *Journal of the Acoustical Society of America* 128.3: 1322–1336.

Gregory, Stanford, Jr., and Brian Hoyt. 1982. Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psycholinguistic Research* 11.1: 35–46.

Gregory, Stanford, Jr., and Stephen Webster. 1996. A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perception. *Journal of Personality and Social Psychology* 70.6: 1231–1240.

Hazan, Valerie, and Stuart Rosen. 1991. Individual variability in the perception of cues to place contrasts in initial stops. *Perception and Psychophysics* 49.2: 187–200.

Ireland, Molly, Richard Slatcher, Paul Eastwick, Lauren Scissors, Eli Finkel, and James Pennebaker. 2011. Language style matching predicts relationship initiation and stability. *Psychological Science* 22.1: 39–44.

Jurafsky, Daniel, Alan Bell, Michelle Gregory, and William Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In *Frequency and the Emergence of Linguistic Structure*, ed. J. Bybee and P. Hopper, 229–254. Amsterdam: John Benjamins.

Kim, Midam, William Horton, and Ann Bradlow. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor distance. *Journal of Language and Politics* 2: 125–156.

Labov, William. 2006. A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics* 34: 500–515.

Levitan, Rivka, Agustín Gravano, Laura Willson, Štefan Beňuš, Julia Hirschberg, and Ani Nenkova. 2012. Acoustic-prosodic entrainment and social behavior. *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, ed. NAACL HLT '12, 11–19.

Lewandowski, Natalie. 2012. Automaticity and consciousness in phonetic convergence. Paper presented at the 2012 Meeting of the Listening Talker Workshop, Edinburgh.
McIntosh, Daniel, Daniel Druckman, and Robert Zajonc. 1994. Socially induced affect. In *Learning, Remembering, Believing: Enhancing Human Performance*, ed. D. Druckman and R.A. Bjork, 251–276. Washington, DC: National Academy Press.
Natale, Michael. 1975. Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills* 40: 827–830.
Nelson, Thomas, and Louis Narens. 1980. Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior* 19: 338–368.
Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39: 132–142.
Nilsenová, Marie, and Palesa Noltig. 2010. Linguistic adaptation in semi-natural dialogues: Age comparison. *Text, Speech and Dialogue, Lecture Notes in Computer Science* 6231: 531–538.
Nguyen, Noël, Sophie Dufour, and Angèle Brunellière. 2012. Does imitation facilitate word recognition in a non-native regional accent? *Frontiers in Psychology* 3: article 480.
Pardo, Jennifer. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119: 2382–2393.
Pardo, Jennifer. 2010. Conversational role influences speech imitation. *Attention, Perception, and Psychophysics* 72.8: 2254–2264.
Pardo, Jennifer, Adelya Urmanche, Sherilyn Wilman, and Jaclyn Wiener. 2015. Phonetic convergence across multiple measures and model talkers. Paper presented at the 2015 Annual Meeting of the Psychonomic Society, Chicago.
Putman, William, and Richard Street, Jr. 1984. The conception and perception of noncontent speech performance: implications for speech-accommodation theory. *International Journal of the Sociology of Language* 46: 97–114.
Pierrehumbert, Janet. 2002. Word-specific phonetics. In *Laboratory Phonology VII*, ed. C. Gussenhoven and N. Warner, 101–140. Berlin: Mouton de Gruyter.
Street, Richard, Jr. 1984. Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research* 11.2: 139–169.
Street, Richard, Jr., and Howard Giles. 1982. Speech Accommodation Theory: A social cognitive approach to language and speech behavior. In *Social Cognition and Communication*, ed. M. Roloff and C. Berger, 193–226. Beverly Hills: Sage Publications.
Yu, Alan. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. In *Origins of Sound Change: Approaches to Phonologization*, ed. A. Yu, 201–227. Oxford: Oxford University Press.

Department of Linguistics
18 Seminary Place
Rutgers University
New Brunswick, NJ 08901
chelsea.sanker@rutgers.edu