UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Distinct behaviors in convergence across measures

Permalink

https://escholarship.org/uc/item/3g39j2zc

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors

Priva, Uriel Cohen

Sanker, Chelsea

Publication Date

2018

Distinct behaviors in convergence across measures

Uriel Cohen Priva[*] (`uriel_cohen_priva@brown.edu`)
Brown University
Providence, RI 02912 USA

Chelsea Sanker[*] (`cas443@cornell.edu`)
Brown University
Providence, RI 02912 USA

Abstract

We present data on convergence in the Switchboard corpus, addressing differences across measures and across speakers. We measured convergence in four characteristics, to test consistency in related and unrelated measures: F0 median, F0 variance, speech rate, and odds of the fillers *uh* and *um*. Convergence was significant in all measures and exhibited variation both between individuals and within individuals. Most notably, convergence in one measure was not predictive of convergence in other measures, except between closely related measures. The results demonstrate some of the limitations of generalizing convergence results from one measure to other measures.

keywords: convergence, individual differences, pitch, speech rate, fillers

Introduction

Speakers' linguistic patterns are influenced by their recent input; their characteristics tend to shift towards being more similar to speech of interlocutors, in a phenomenon called *convergence*, among other names (cf. Giles, Taylor, & Bourhis, 1973; Goldinger, 1998). In a study of the Switchboard corpus (Godfrey & Holliman, 1997), we demonstrate that while convergence is observable across multiple measures, there are differences between measures, and that individual variation in convergence is not consistent across measures.

Convergence has been observed in many behaviors, including non-linguistic behaviors such as fidgeting (Chartrand & Bargh, 1999) and posture (Dijksterhuis & Bargh, 2001) and also many linguistic characteristics, including vowel formants (e.g. Babel, 2012), VOT (e.g. Nielsen, 2011), F0 (e.g. Babel & Bulatov, 2011), jitter and shimmer (e.g. Levitan et al., 2012), lexical choice (e.g. Branigan, Pickering, Pearson, McLean, & Brown, 2011), syntactic constructions (e.g. Branigan, Pickering, & Cleland, 2000), and timing of turns and pauses (e.g. Natale, 1975; Street, 1984), among others. For many measures, convergence has been demonstrated both at the conversation level (e.g. Levitan et al., 2012; Nielsen, 2011) and dynamically within conversations (e.g. Street, 1984; Vaughan, 2011). Convergent shifts can persist even after the end of the conversation or exposure to stimuli (e.g. Babel, 2012; Pardo, 2006).

Convergence is also apparent in holistic measures based on listeners' decisions about the similarity of participants' utterances before and after a task to the utterances of the model talker (e.g. Goldinger, 1998; Pardo, 2006). There is some evidence for a correlation between results from holistic perceptual measures of convergence and acoustic characteristics (Pardo, 2009), though other studies have found a lack of correlation between perceptual results and acoustic results (Babel & Bulatov, 2011; Pardo, Gibbons, Suppes, & Krauss, 2012).

Despite the existence of much data demonstrating that convergence occurs in many characteristics, there is little work examining whether or not convergence is comparable across different characteristics. Much work only measures convergence in one characteristic, based on its relevance for a particular phenomenon or theory, e.g. effects of socially salient dialect differences (Drager, Hay, & Walker, 2010) or generalization across phonological features (e.g. Nielsen, 2011). Other work includes multiple measures but does not directly compare the results across characteristics, instead focusing on testing factors that can influence degree of convergence, such as race (Babel, 2012), gender (Bilous & Krauss, 1988; Pardo, Jay, & Krauss, 2010), and status (Gregory & Webster, 1996). Such work usually does not explicitly address how the choice of measure might influence the result.

Different measures of convergence have been demonstrated to correlate with some of the same social factors, such as social characteristics of the speaker (Natale, 1975), partners' ratings of their relationship (Pardo et al., 2012), and observers' ratings of the partners' relationship (Levitan et al., 2012). Some convergence measures have also been observed to correlate similarly with the same objective measures, such as amount of overlapping speech (Levitan et al., 2012).

However, there are few direct comparisons across measures. There is substantial variation in how each characteristic is influenced by aspects of the task and the participants, such as word frequency and speaker gender (e.g. Bilous & Krauss, 1988; Pardo, Urmanche, Wilman, & Wiener, 2017), as well as in the degree of convergence exhibited by different characteristics, both unrelated characteristics, e.g. formants, F0, and turn durations (Sanker, 2015) and related characteristics, e.g. formants of different vowels (Babel, 2012).

There is even less data on individual speakers' tendency to converge. While previous studies have found variation in convergence across participants (e.g. Babel, 2012; Pardo et al., 2010), it is not clear how much of the variation reflects consistent characteristics of particular individuals and whether these patterns would be present in the same speakers in different tasks or using different measures of convergence.

---

[*]Both authors contributed equally to this manuscript.

Tamminga, Wade, and Lai (2018) found consistency in individual speakers' patterns of convergence in two instances of the same task. Sanker (2015) similarly found consistency within the same interlocutor pair in different conversational tasks and in the same speaker in different pairs, but no tendency for individual consistency across different measures. Bilous and Krauss (1988) also found differences in individual pairs' degree of convergence in different measures. On the other hand, Rahimi, Kumar, Litman, Paletz, and Yu (2017) found a trend for positive correlations of convergence in some different measures, though it was not consistent across different comparisons.

Some scholars have identified certain personality traits that partially predict differences in convergence (Chartrand & Bargh, 1999; Natale, 1975; Yu, Abrego-Collier, & Sonderegger, 2013), and argue that individual differences in attention to detail make some individuals more likely to exhibit convergence in laboratory studies and also, on a larger scale, to propagate sound change (Yu et al., 2013). This model of individual difference suggests that different phonetic characteristics are likely to behave the same way, which is not clearly reflected in existing data. While such studies offer correlations of convergence or performance in other tasks with these external measures, they have not demonstrated that individual variation is consistent across tasks.

In this paper, we examine variation in convergence across speakers and across measures. Beyond demonstrating that both types of variation exist, we show that the patterns exhibited by each speaker in one measure are not predictive of that speaker's patterns in other measures. To the best of our knowledge, this study is the first large-scale investigation of variation in convergence both across measures and across speakers. It is also the first study to examine individual tendencies in convergence with a method that controls for effects that could be due to particular conversations and not convergence per se.

Methods

Corpus

The data for this study is the Switchboard Corpus (Godfrey & Holliman, 1997), a large collection of telephone conversations. Each speaker was randomly paired with other speakers and given a topic for each conversation, providing a large corpus of natural speech data for many speakers in similar conversations with several different partners; recordings have caller identification information that can be used to compare the conversation with other instances of that caller. Each side of the conversation is a distinct recording, so measurements can reliably be taken for each speaker separately.

Each conversation has associated information quantifying the clarity of the recording; after omitting calls with high levels of background noise, echoing, or other issues, as indicated in the annotations of these calls, the set of data that we used had 464 speakers, in 3782 conversations.

Speech rate measures were based on the manually corrected word annotations produced at MS State (Harkins, Feinstein, Lindsey, Martin, & Winter, 2003), which allow measurements of word duration.

Measures Used

The measures used were selected to provide a range of speech characteristics, both related and unrelated, to compare convergence patterns in different types of measures. The methods for calculating each measure are given below; the F0 measures differ in some ways from more common versions of these measures established in prior work, in order to minimize errors due to automated measurements in a large corpus.

F0 median: Measured in Mels (Stevens, Volkmann, & Newman, 1937) and excluding tokens beyond density minima at either end of the distribution, to exclude data points due to pitch tracking errors. While this might also exclude some cases of actual extreme F0s, checking some of the apparent outliers confirmed that most of them are the result of erroneous pitch tracking halving or doubling the actual F0.

F0 range: Log of the ratio of the 75th percentile to 25th percentile of F0 measurements in Hertz. Using the quotient rather than the difference was aimed at reducing the artificial correlation between F0 median and F0 range; with a difference or a standard deviation, the range would scale up in proportion to the center of the distribution in a way that does not align with listener perceptions (cf. Jessen, Köster, & Gfroerer, 2005; Stevens et al., 1937). Using the quartiles reduced the sensitivity of the measurements to outliers and the outlier exclusion method employed in the F0 data.

Speech rate: Measured relative to the predicted duration based on each word's median duration median duration within the Switchboard corpus, the length of the utterance, and the distance from the end of the utterance. See Cohen Priva, Edelist, and Gleason (2017) and Cohen Priva (2017) for a discussion of the benefits of measuring speech rate in this way rather than as a raw value.

Uh-Um log odds: The relative odds of encountering each of the fillers *uh* and *um* in the speech of each talker (cf. Acton, 2011; see also Clark & Fox Tree, 2002). To calculate log odds, a logistic regression was performed for all pairs of counts (e.g. <20 *um*, 2 *uh*>), and the log odds ratio was the predicted value of the regression plus the residuals of the regression. The advantage of using this method is that it can produce log odds for speakers who never used one or the other. Usage of fillers has been observed to vary, and some usage patterns align with interspeaker differences (Acton, 2011).

Measuring Convergence

Convergence was measured by comparing how speakers' productions within a conversation differ from their baselines in the direction of their partner's baselines, measured from all conversations except for the one with the partner under consideration, i.e. establishing independent baselines for each speaker and looking at the degree to which partners are converge to the baselines of their interlocutors (cf. Cohen Priva

et al., 2017). This method is aimed at removing influences due to the conversation rather than the interlocutor.

Many convergence studies compare speakers' productions within a conversation to their interlocutor's productions within that conversation. Using this method, situations in which both speakers shift in the same way will appear to be convergent, even though increase in similarity within a conversation can have a variety of causes that do not depend on sensitivity to the interlocutor's speech, e.g. effects of the conversational topic or task.

Some work includes comparisons with other speakers performing the same task, to control for task-related effects (e.g. Levitan & Hirschberg, 2011; Sanker, 2015), but this does not control for effects of the particular conversation. There is significantly greater similarity of speakers to their interlocutors as compared within a conversation than as compared to interlocutors' characteristics from other conversations (Gregory & Webster, 1996). While this method may decrease the amount of actual convergence captured, convergence is still apparent when tested in this way.

Establishing reliable baselines depends on having a large corpus, so that baselines are averaged across enough conversations to not be thrown off by the particular characteristics of any particular conversation.

Statistical Models

Data was analyzed in R (R Core Team, 2017) with mixed effects models. There were separate models with each of the linguistic measures as the variable, for a total of four models. All models had two fixed effects: (1) the mean of the speaker's performance in other conversations, and (2) the mean of the interlocutor's performance in other conversations. Thus, strong *consistency* across conversations would be reflected in high coefficient values for the speaker's performance in other conversations, and strong *convergence* would be reflected in high coefficient values for the interlocutor's performance in other conversations. Speaker identity was not used as a random intercept, because characteristic patterns of individual speakers are better modeled by their respective baselines and models including both factors might fail to converge due to high collinearity. The models included a random slope for the interlocutor's baseline performance, which was used to model the different degrees of convergence different that speakers may exhibit. The coefficients provided below were modeled using the lmerTest package (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2015) which builds on lme4 (Bates, Mächler, Bolker, & Walker, 2015) to include a calculation of degrees of freedom and p-values.

Models produced by lme4 returned zero variance for several of the random slopes for speaker, which seemed to be an unlikely estimate of individual speakers' consistency in convergence across conversations. We therefore retrained the models using the brms package (Bürkner, 2017), which found more non-zero random slopes, but produced less consistent results due to its sampling-based nature. In order to estab-

lish a more consistent estimate of individual variation, we repeated the sampling procedure 350 times for each model and used the median value of per-speaker estimates. These values are used below when testing correlations between speakers' convergence in different domains (cf. Tamminga, 2017).

Results

Speaker and interlocutor baselines as predictors

Within the mixed effects models for each speech characteristic, by far the main predictor of that variable is the speaker's mean performance in that measure from other conversations, which was highly significant for all of the four measures investigated, as given in Table 1. That is, speakers were very consistent in their production patterns across conversations.

Table 1: Speaker baseline as a predictor of each variable

	β	SE	t value	p value
F0 median	0.971	0.00387	250.71	< 0.0001
F0 var.	0.676	0.012	56.39	< 0.0001
log(uh:um)	0.788	0.009	87.65	< 0.0001
speech rate	0.795	0.0088	90.37	< 0.0001

Interlocutor baseline as a predictor of each variable was also significant, though the effect was much smaller than speaker baseline. This measure was capturing convergence; the positive coefficient in all cases, as given in Table 2, indicates convergence in all measures.

Table 2: Interlocutor baseline as a predictor of each variable

	β	SE	t value	p value
F0 median	0.0176	0.00404	4.36	< 0.0001
F0 var.	0.0924	0.0124	7.47	< 0.0001
log(uh:um)	0.0311	0.0099	3.14	0.00186
speech rate	0.0471	0.0088	5.35	< 0.0001

Some characteristics exhibited more convergence than others, but all measures exhibited significant convergence and the size of the effect was within the same order of magnitude. The measure which exhibited the strongest evidence of convergence was F0 variability.

Correlations between measures, by individual

The individual-level variation between speakers in degree of convergence, i.e. the extent to which their productions were predicted by the interlocutor's baseline in a particular measure, would be close to zero if speakers were not consistent in the degree to which they converged across conversations. Consistency of convergent behavior within a speaker would be reflected in non-zero standard deviation for the random slope in each model. The models generated by the brms package consistently resulted in standard deviation estimates that were positive and of the same order of magnitude as the coefficient for the interlocutor's baseline performance; however,

the 95% confidence interval included 0 for all models except the *uh* to *um* ratio model. This result indicates that speakers' degree of convergence in one conversation was only a weak predictor of their convergence in other conversations.

However, differences between individual speakers' convergence tendencies were large enough to allow a comparison of convergence between different measures, by speaker. Speakers exhibited little consistency in degree of convergent change across different characteristics, as illustrated in Table 3; a speaker's convergence patterns in one measure were not predictive of that speaker's convergence patterns in other measures.

Table 3: Correlations between speaker-level convergence in each pair of measures (F0 median, F0 range, log odds of *uh*:*um* ratio, and speech rate).

	F0 var.	uh:um	speech rate
F0 median Pearson *r*	0.22	0.07200	-0.0305
Sig. (two-tailed)	< 0.0001	0.12200	0.5120
F0 var. Pearson *r*		0.00027	0.0596
Sig. (two-tailed)		0.99500	0.2000
log(uh:um) Pearson *r*			-0.0702
Sig. (two-tailed)			0.1260

The only comparison in which the correlation in convergence across speakers was significant was between F0 median and F0 variability. To rule out the possibility that the correlation was due to the strong correlation that exists in production between F0 variance and F0 median (Jessen et al., 2005), we fitted a linear regression with cubic functions applied to F0 median in Mels, log F0 median in Hz, and F0 median. We then used the model to extract the *residuals* of F0 variance, i.e. the component of F0 variance that was not explained by the three predictors. We then repeated the procedure outlined above for the F0 variance model, using the residualized values. The two measures were still significantly correlated, albeit to a lesser extent (Pearson *r* = 0.106, *p*= 0.022).

Among pairs of measures other than F0 median and F0 variability, there were no significant correlations in convergence, nor any trend towards positive correlation. Notably, there is no intrinsic correlation in production between any of these pairs of measures. The large number of speakers (n = 464) makes it unlikely that a lack of correlation could result from an inadequate sample size, which would be a concern in a smaller scale study. In addition, the significant correlation between F0 variance and F0 median demonstrates that these methods can capture individuals' consistency across measures when a relationship exists, despite individual speakers having only a weak trend towards consistency in convergence across conversations. Thus, the results are likely to be capturing an actual lack of relationship between convergence in different measures.

Discussion and Conclusion

In convergence studies, the measure used can have a large impact on the results, as different measures can exhibit different overall degrees of convergence as well as different influences (Bilous & Krauss, 1988; Pardo et al., 2017; Sanker, 2015). We extend the data on this variation within a large corpus of natural speech, confirming differences in convergence measured in different characteristics.

The different size of the convergence effect in different measures has potential implications for design of future convergence studies. While all of the measures exhibited significant convergence, the differences are large enough that in a smaller sample, they might not all reach significance, which makes the measure with the most convergence, F0 variability, a promising characteristic to use in measuring convergence, at least within conversational tasks; it is not frequently used, though there are some convergence studies that have included it (e.g. Vaughan, 2011). It may be that measures of variation have a slight advantage in capturing the dynamic aspects of convergence, while means and medians obscure some of it by collapsing over long time spans.

In addition to variation in convergence across measures, there is variation across speakers. However, speakers were not strongly consistent in their degree of convergence across conversations, suggesting that convergence is more influenced by aspects of particular conversations than characteristics of each individual independently. Other work has found a larger effect of speaker consistency, at least within closely related tasks: Between instances of same task, either a conversation with a set topic (Sanker, 2015) or shadowing of set stimuli (Tamminga et al., 2018), or between different conversational tasks with the same partner (Sanker, 2015). Individual tendencies in convergence might be more apparent with more constrained conditions across the tasks being compared, because there is less possibility of an effect of contextual factors like interlocutor and conversational topic.

Convergence exhibited by a speaker in one measure was not correlated with convergence in other measures. The lack of correlation between measures indicates that variation in convergence across speakers cannot be attributed to consistent differences in processing style, with different listeners focusing more or less on low-level detail (cf. Yu, 2013). Rather, the results suggest that individual differences in attention to detail might depend on the particular characteristic, which is consistent with variation across listeners in which acoustic cues they attend to for the same phonological and structural contrasts (e.g. Hazan & Rosen, 1991; Roy, Cole, & Mahrt, 2017). On the other hand, other studies have found a lack of correlation in cue weighting between perception and production, e.g. in participants' use of F0 and VOT as cues for stop contrasts (e.g. Schertz, Cho, Lotto, & Warner, 2015), so it is not clear whether perceptual weighting of different cues would extend to convergence or not. The lack of consistency in convergence across measures might also in part be due to speakers' variability in convergence across conversations.

The one exception to the independent patterns of convergence in different measures was the relationship between F0 median and in F0 range. Adapting the model for F0 variance to include F0 median as a predictor substantially reduced the correlation in convergence, which suggests that this correlation is largely an effect of the correlation in production between F0 median and F0 range. However, the correlation in convergence between these two measures was still significant in this model, which might suggest a perceptual link, with attention to low-level detail in F0 reflected in both measures. On the other hand, it could also be due to an indirect effect of other factors that are correlated with both characteristics, though comparing speakers' productions to their partners' baselines rather than their partners' measurements within their shared conversations makes this less likely.

Phonetic convergence is often presented as evidence for episodic memory of utterances, in which speakers store details of each instance of hearing a word or phoneme, with greater weight given to recent exemplars (e.g. Goldinger, 1998). Though such models do not specifically address predictions about variation across speakers and across linguistic characteristics, the observed differences in convergent behavior could be consistent with a hybrid exemplar model in which exemplar clouds are shaped by a system of abstractions (e.g. Pierrehumbert, 2002), such that speakers can differ in how they weight exemplars for different characteristics. This different weighting of cues could easily be integrated into the model, as it already allows differential weighting of recent and otherwise salient exemplars.

While a shift in representation based on episodic memory is strongly supported by convergent shifts which continue at long delays after input, this effect has not been tested for all measures, and they might not all behave similarly. Some convergent effects may be based on priming and perceptual-behavioral links rather than a shift in representation, as is proposed by some analyses, particularly for non-linguistic convergence (e.g. Dijksterhuis & Bargh, 2001; Giles et al., 1973). Characteristics which are cues to a phonological contrast, such as vowel formants, might also have different representations than characteristics which are not associated with a phonological contrast, such as F0 in English. The existence of multiple explanations underlying convergence would be consistent with different convergence patterns in different measures; comparisons across measures can help test the predictions made by different explanations and representations.

The differences in convergence in different characteristics demonstrate the importance of considering convergence separately for different measures, not just in building linguistic models but also when interpreting experimental results, as convergence patterns observed in one characteristic might not be paralleled in other characteristics. In addition, the lack of correlation in individuals' behavior across measures demonstrates a limitation of using individual variation in results from a single measure to characterize individual differences in perception or phonological processing.

References

Acton, E. K. (2011). On gender differences in the distribution of um and uh. In *University of Pennsylvania working papers in linguistics* (Vol. 17, chap. 2).

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

Babel, M., & Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, *55*(2), 231–248. doi: 10.1177/0023830911417695

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bilous, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviours of same-and mixed-gender dyads. *Language & Communication*, *8*(3), 183–194. doi: 10.1016/0271-5309(88)90016-X

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*, B13–25. doi: 10.1016/S0010-0277(99)00081-5

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*, 41–57. doi: 10.1016/j.cognition.2011.05.011

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: 10.18637/jss.v080.i01

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893. doi: 10.1037/0022-3514.76.6.893

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. doi: 10.1016/S0010-0277(02)00017-3

Cohen Priva, U. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition*, *160*, 27–34. doi: 10.1016/j.cognition.2016.12.002

Cohen Priva, U., Edelist, L., & Gleason, E. (2017). Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *Journal of the Acoustical Society of America*, *141*(5), 2989–2996. doi: 10.1121/1.4982199

Dijksterhuis, A., & Bargh, J. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology*, *33*, 1–40. doi: 10.1016/S0065-2601(01)80003-4

Drager, K., Hay, J., & Walker, A. (2010). Pronounced rivalries: Attitudes and speech production. *Te Reo*, *53*, 27–53.

Giles, H., Taylor, D., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society*, *2*, 177–192.

Godfrey, J. J., & Holliman, E. (1997). *Switchboard-1 release*

2. (Linguistic Data Consortium, Philadelphia)

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. doi: 10.1037/0033-295X.105.2.251

Gregory, S. W. J., & 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.

Harkins, D., Feinstein, D., Lindsey, T., Martin, S., & Winter, G. (2003). *Switchboard MS State manually corrected word alignments.* https://www.isip.piconepress.com/projects/switchboard/.

Hazan, V., & Rosen, S. (1991). Individual variability in the perception of cues to place contrasts in initial stops. *Perception & Psychophysics*, *49*(2), 187–200. doi: 10.3758/BF03205038

Jessen, M., Köster, O., & Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language and the Law*, *12*(2), 174–213.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2015). lmerTest: Tests in linear mixed effects models [Computer software manual]. (R package version 2.0-29)

Levitan, R., Gravano, A., Willson, L., Beňuš, Š., Hirschberg, J., & Nenkova, A. (2012). Acoustic-prosodic entrainment and social behavior. In *Proceedings of the conference of the North American chapter of the Association for Computational Linguistics: Human language technologies* (pp. 11–19). Stroudsburg: Association for Computational Linguistics.

Levitan, R., & Hirschberg, J. B. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of Interspeech.* Brisbane: International Speech Communications Association.

Natale, M. (1975). Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills*, *40*, 827–830.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*, 132–142. doi: 10.1016/j.wocn.2010.12.007

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

Pardo, J. S. (2009). Expressing oneself in conversational interaction. In E. Morsella (Ed.), *Expressing oneself/expressing one's self: Communication, cognition, language, and identity* (pp. 183–196). Hove: Psychology Press.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*, 190–197. doi: 10.1016/j.wocn.2011.10.001

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversa-

tional role influences speech imitation. *Attention, Perception, and Psychophysics*, *72*(8), 2254–2264. doi: 10.3758/BF03196699

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

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101–140). Berlin: Mouton de Gruyter.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna.

Rahimi, Z., Kumar, A., Litman, D., Paletz, S., & Yu, M. (2017). Entrainment in multi-party spoken dialogues at multiple linguistic levels. In *Proceedings of Interspeech* (pp. 1696–1700). Stockholm, Sweden.

Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Laboratory Phonology*, *8*(1), 1–36. doi: 10.5334/labphon.108

Sanker, C. (2015). Comparison of phonetic convergence in multiple measures. In *Cornell working papers in phonetics and phonology 2015* (pp. 60–75).

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183–204. doi: 10.1016/j.wocn.2015.07.003

Stevens, S., Volkmann, J., & Newman, E. (1937). A scale for the measurement of psychological magnitude pitch. *Journal of the Acoustical Society of America*, *8*, 185–190.

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, *11*(2), 139–169. doi: 10.1111/j.1468-2958.1984.tb00043.x

Tamminga, M. (2017, May). *How much do individuals vary?* Talk presented at the Stanford Linguistics Department's colloquium.

Tamminga, M., Wade, L., & Lai, W. (2018, January). *Stability and variability in phonetic flexibility.* Talk presented at the Linguistic Society of America annual meeting. (https://www.linguisticsociety.org/abstract/stability-and-variability-phonetic-flexibility)

Vaughan, B. (2011). Prosodic synchrony in co-operative task-based dialogues: A measure of agreement and disagreement. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of Interspeech* (pp. 1865–1868). ISCA.

Yu, A. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. In A. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 201–227). Oxford, England: Oxford University Press.

Yu, A., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *Journal of Statistical Software*, *8*(9), e74746. doi: 10.1371/journal.pone.0074746