Dimensions of Convergence

Chelsea Sanker

14 February 2020

Phonetics & Experimental Phonology Laboratory
New York University

Introduction

- Convergence is the phenomenon in which speakers' productions become more similar to their interlocutors'
- What are the dimensions that shape variability and consistency in convergence (how does it generalize)?
  - Individual speakers and interlocutors
  - Particular words
  - Particular sounds or features
  - The characteristic being measured

Individual tendencies in convergence

- Degree of convergence is correlated with some characteristics of individuals (Natale 1975; Yu et al. 2013), attributed to individual cognitive differences that produce convergent tendencies
- Evidence for individual convergent tendencies exists within the same task or similar tasks (Sanker 2015; Tamminga et al. 2018), weaker across distinct tasks (Pardo et al. 2018)
- Little evidence for individual tendencies that hold across characteristics (Bilous and Krauss 1988; Pardo et al. 2012; Sanker 2015; Weise and Levitan 2018)

Individual tendencies in convergence

Much work includes each participant in a single interaction, making it impossible to distinguish effects of individuals vs. the particular conversation or of an individual as a speaker vs. an interlocutor

- Interlocutor/model talker effects: Differences in degree of convergence elicited by different model talkers (e.g. Pardo et al. 2017; Hwang and Chun 2018)
- Conversation effects: Differences based on the particular pair of interlocutors and their relationship (e.g. Pardo et al. 2012; Bane et al. 2010; Sanker 2015)

Convergence measured in different characteristics

If there are individual tendencies in convergence, are they apparent across different characteristics?

Convergence has been observed in a range of characteristics, e.g.

- vowel formants (e.g. Babel 2012)
- VOT (e.g. Nielsen 2011)
- f0 (e.g. Babel and Bulatov 2011)
- speech rate (e.g. Cohen Priva, Edelist, and Gleason 2017)
- holistic impressions of speech similarity (e.g. Goldinger 1998)

But most work finds little consistency across measures, either by conversation or by speaker

Convergence across words

Can convergence be lexically specific, providing evidence for lexically-specific phonetic representations?

- Extension across words (Pardo et al. 2012; Nielsen 2011) – But in previous work the shifted characteristic in training was consistent across words
- Lexical frequency effects – more convergence in lower frequency words, perhaps suggesting lexical specificity (Goldinger 1998; Babel 2010; Nielsen 2011; Dias & Rosenblum 2016), but:
  - Other studies have failed to replicate the effect (Pardo et al. 2013; Pardo et al. 2017)
  - All studies that found a frequency effect used a single model talker
  - Different behaviors based on lexical frequency are an indirect test of word-specific convergence

Convergence across sounds

What are the targets of convergence? What characteristics generalize and what categories do they generalize across?

- Using a shadowing task, Nielsen 2011 demonstrates extension of lengthened VOT in /p/ to VOT in /k/
- Similar work in perceptual training finds extension for some things (e.g. VOT across places of articulation, Kraljic and Samuel 2006; fricative place across voicing category, Schuhmann 2014), but not others (e.g. place of articulation across stops and nasals, Reinisch et al. 2014)

Study 1a: Questions about individual tendencies

This study is collaborative work with Uriel Cohen Priva

Many studies find variation in convergence across participants. Can it be attributed to individual tendencies for convergence?

- Is there consistency in convergence *produced* by an individual across conversations with different partners?
- Is there consistency in convergence *elicited* by an individual across conversations with different partners?
- Is there consistency across measures within a conversation?

Methods

Using the Switchboard corpus:

- Short phone conversations, randomly paired with strangers and given a topic to discuss
- ∼2000 conversations between pairs of speakers
- ∼520 unique speakers
- Hand-corrected word-level alignment
- Analyzed in Praat

Methods

Characteristics (all were normalized to facilitate comparison):

- F0 median
- F0 range
- speech rate
- lexical information rate (negative log probability of words)
- filled pause type (log odds of uh:um)
- sentence-initial 'and' (log odds with and without it)

Methods

Measuring convergence with mixed effects linear regression models, with the speaker's measured productions in each characteristic as the dependent variable.

Fixed effects

- The speaker's production of that characteristic in other conversations (*consistency*)
- The interlocutor's production of that characteristic in other conversations (*convergence*)

Random effects

- Intercepts for interlocutor, conversation, topic
- Slopes for convergence by speaker, interlocutor, and conversation
- And per-characteristic versions of each

Results: Fixed effects

The speaker's own baseline was a significant predictor of that speaker's performance (β=0.711, SE=0.079, df=5, t=8.95, p=0.00029): Speakers' productions were highly consistent across conversations

The interlocutor's baseline was also a significant predictor of the speaker's performance (β=0.0471, SE=0.011, df=5, t=4.45, p=0.00596): There was significant convergence

Results: Random effects

Table: Summary of random effects

	SD	Model comp. p
Per-char conversation intercept	0.247	$<$0.0001
Per-char interlocutor slope for convergence	0.0143	0.866
Per-char interlocutor intercept	0.0715	0.00053
Per-char speaker slope for convergence	0.046	0.0442
Per-char speaker intercept	0.000	1.000
Conversation slope for convergence	0.000032	1.000
Interlocutor slope for convergence	0.0454	0.00160
Speaker slope for convergence	0.000	1.000
Per-char topic intercept	0.201	$<$0.0001
Per-char convergence	0.0227	0.00123
Per-char consistency	0.194	$<$0.0001

Study 1a Results Summary

- Some weak evidence for individual tendencies within a characteristic, but nothing across characteristics to suggest a cognitive trait which predisposes some speakers to convergence more than others
- Consistency in convergence by interlocutor, suggesting individual social effects: Some speakers are viewed more positively by conversational partners, which increases convergence
- No evidence for an effect of the particular conversation

Study 1b: Questions about methods of measuring convergence

This study is collaborative work with Uriel Cohen Priva (Cohen Priva and Sanker 2019)

- Can some of the previous results finding apparently robust speaker effects be an artifact of how convergence is measured?
- The difference-in-difference method often used for measuring convergence has several potential issues:
  - Proximity in the baselines of the speaker and interlocutor or model talker may lead to measured divergence (Random variation is likely to produce second measurements that are less similar)
  - Extreme baselines may lead to overestimation of convergence (Regression to the mean)
  - Such artifacts would produce the appearance of individual variation in convergent tendencies

Methods

Using the Switchboard corpus as in Study 1a:

- Short phone conversations, randomly paired with strangers and given a topic to discuss
- ∼2000 conversations between pairs of speakers
- ∼520 unique speakers
- Hand-corrected word-level alignment
- Analyzed in Praat

Methods

Characteristics (all were normalized to facilitate comparison):

- F0 median
- F0 range
- speech rate
- lexical information rate (negative log probability of words)

Methods

The measurements were analyzed in two different ways (both modelled with mixed effects linear regression):

1. Difference-in-difference (DID), the difference between the distance from the reference value (interlocutor's baseline) using the speaker's productions in the shared conversation and in other conversations:
   $|\ ParticipantOtherConvos - Reference\ |$ -
   $|\ ParticipantSharedConvo - Reference\ |$
   - Frequently used in convergence studies
   - DID was the dependent variable (the intercept is the measure of convergence)
2. Linear combination, predicting each speaker's productions based on their productions elsewhere and their interlocutor's productions
   - Not in common use, but makes it possible to account for noise that is not due to convergence
   - The speaker's value in the shared conversation was the dependent variable (the fixed effect of interlocutor is the measure of convergence)

Results: Effect of starting distance in DID models

Coefficients for the absolute distance between the subject's baseline and interlocutor's baseline as a predictor of DID; significant positive values in all models (more convergence with greater starting distance)

	β	SE	df	t	p
F0 median	0.16	0.02	3600	8.2	$<$0.0001
F0 variance	0.54	0.02	3368	29.9	$<$0.0001
Speech rate	0.45	0.02	3604	29.1	$<$0.0001
uh:um ratio	0.43	0.02	4071	25.8	$<$0.0001

Results: Effect of starting distance in linear combination models

Coefficients for the absolute distance between the subject's baseline interlocutor's baseline performance as a predictor of the subjects' performance in linear combination models; not significant in any model

	β	SE	df	t	p
F0 median	-0.0061	0.007	2882	-0.9	0.39
F0 variance	-0.0108	0.013	1215	-0.8	0.42
Speech rate	0.0018	0.008	780	0.2	0.82
uh:um ratio	-0.0138	0.011	1988	-1.2	0.21

Results: Effect of extreme baselines in DID

Coefficients for absolute distance between the subject's baseline performance and the nearest mode of the of the distribution as a predictor of DID; significant positive values in all models (more convergence with greater starting distance)

	β	SE	df	t	p
F0 median	0.166	0.06	3674	2.6	0.00938
F0 variance	0.151	0.04	3660	3.9	0.00012
Speech rate	0.205	0.04	4720	5.1	$<$0.0001
uh:um ratio	0.094	0.03	4667	3.2	0.00120

Results: Effect of extreme baselines in linear combination

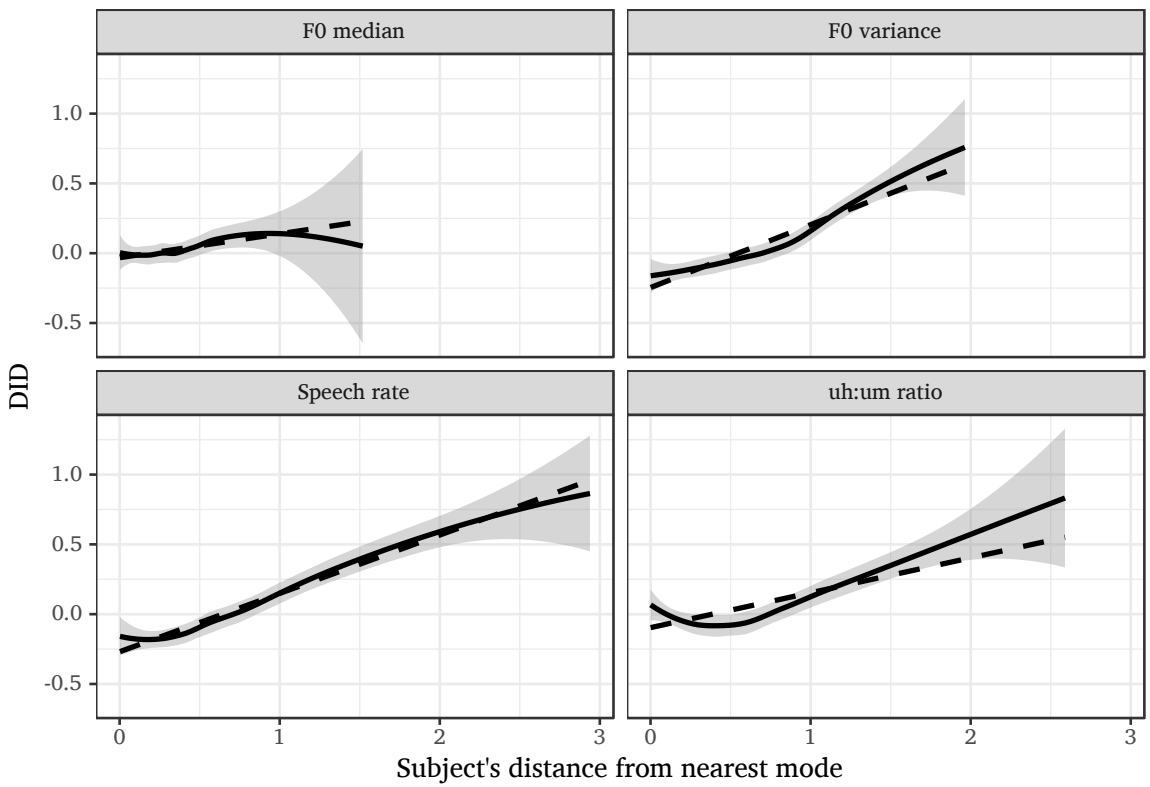
Coefficients for absolute distance between the subject's baseline performance and the nearest mode of the of the distribution as a predictor of the subjects' performance in linear combination models; not significant in any model

	β	SE	df	t	p
F0 median	-0.0085	0.01	266	-0.8	0.41
F0 variance	0.0078	0.02	371	0.4	0.72
Speech rate	0.0122	0.01	286	0.8	0.40
uh:um ratio	0.0113	0.02	375	0.6	0.52

Results: Individual differences using DID

Individual differences in convergence in DID models were measured as the mean per-conversation DID values for each subject.

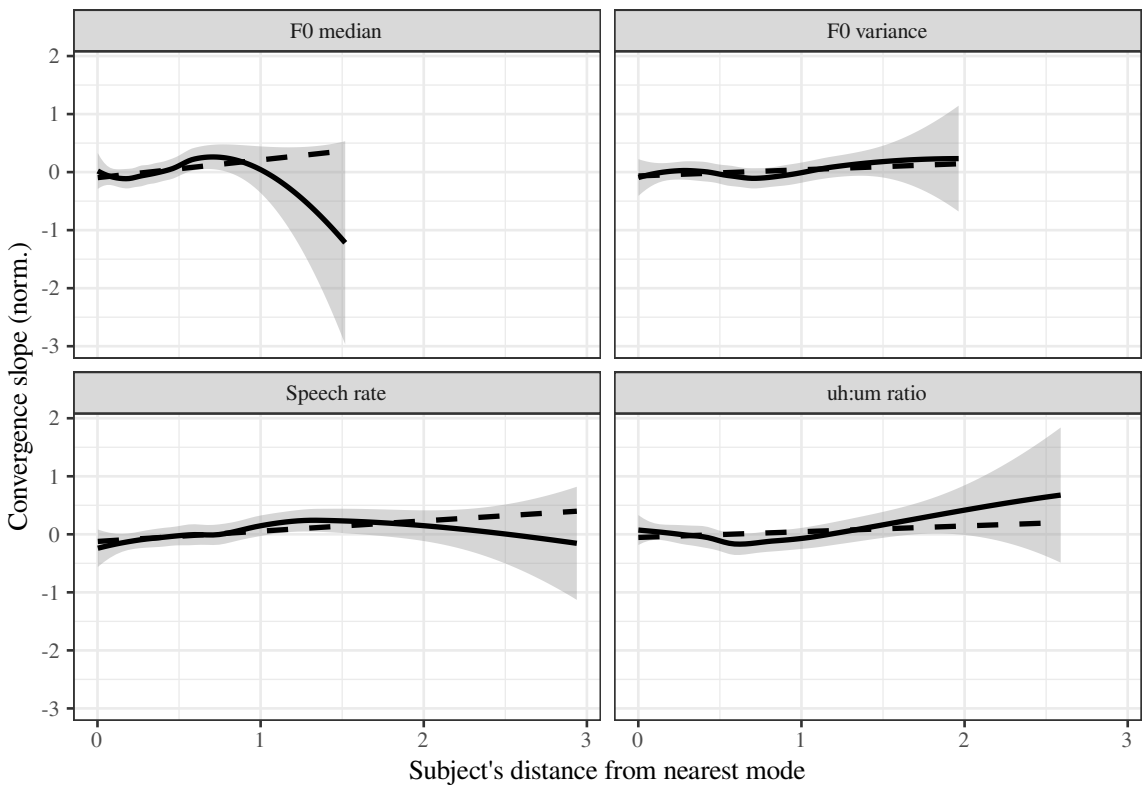
The subject's distance from the nearest mode predicted measured convergence in all characteristics.



Results: Individual differences using linear combination

Individual differences in convergence in linear combination models were measured as the per-subject random slope for the interlocutor's baseline.

There was little relationship between distance from the nearest mode and measured convergence.



Follow-up: Simulated data

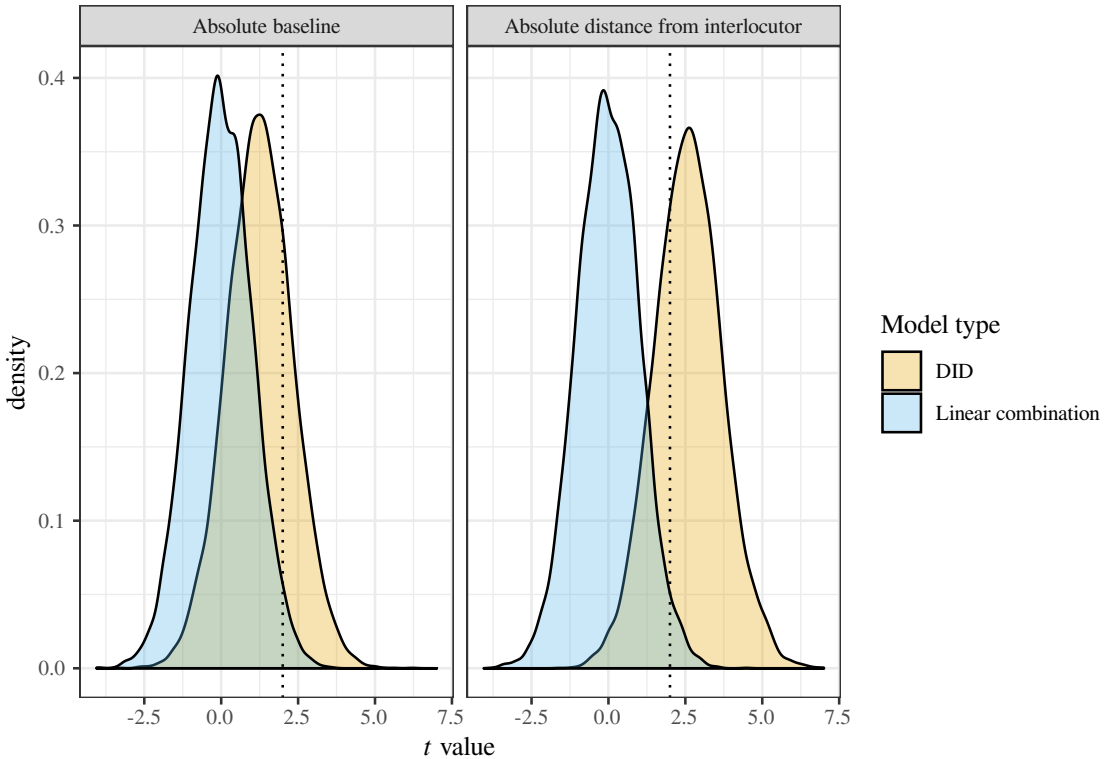
Simulated data was generated for 50 pairs of speakers. The true baseline of each participant was sampled from a normal distribution of the population, with a mean of 0 and a standard deviation of 1.

The *before* and *after* performances for each participant were calculated from the participant's baseline value with normally distributed noise with a mean of 0 and a standard deviation of 0.5.

The participants' productions were compared to the productions of another randomly sampled participant.

That is, there was some variation in each speaker's productions (self correlation $= 0.8$), but there was no convergence.

Follow-up: Simulated data

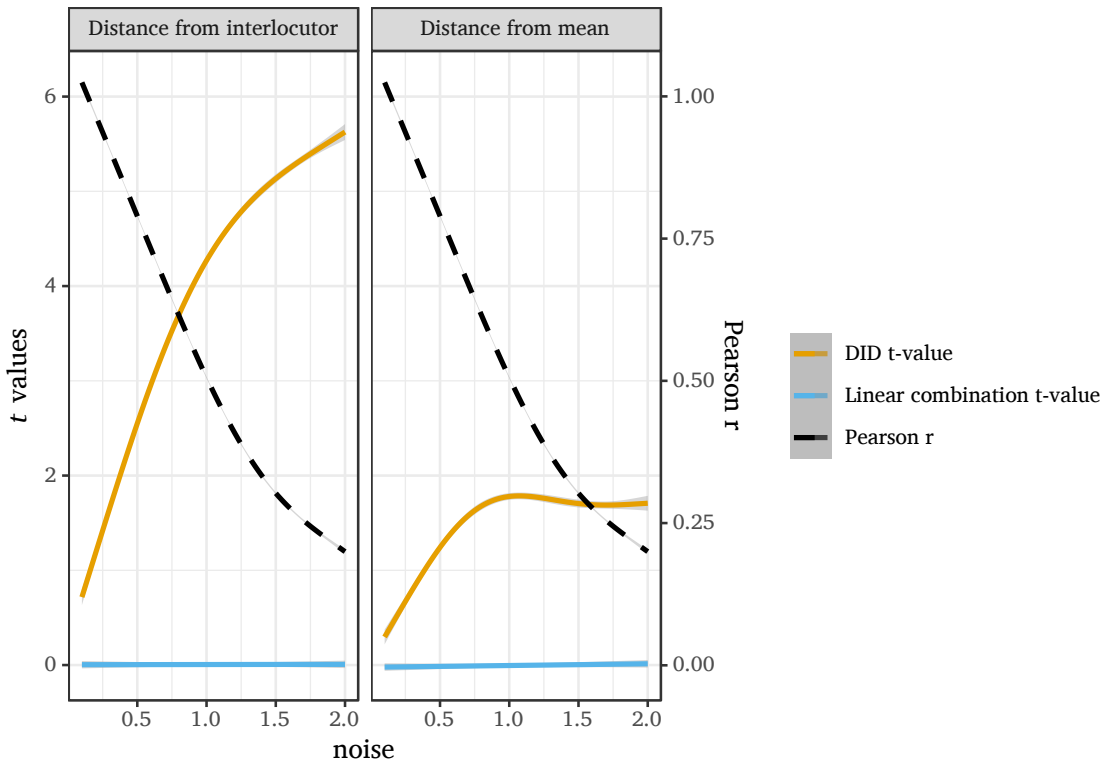


Density plots of *t*-values for distance from the population mean and baseline distance from the interlocutor in 10,000 samples.

In DID models, the subjects' distance from the mean had significant positive coefficients 24.7% of the time, and distance between the speaker and interlocutor 70.8% of the time.

In linear combination models, both were significant about 2.5% of the time.

Follow-up: Simulated data



The correlations found by DID models increase when there is more noise in the data. Even in very noisy data, the correlations do not appear in linear combination models.

Study 1b Results Summary

- Two artifacts of the difference-in-difference method are apparent:
  - Close proximity between the speaker and interlocutor/model leads to measured divergence or underestimated convergence
  - Extreme baselines produce overestimated convergence
- These artifacts are present in real data and confirmed with simulated data defined to lack convergence
- DID is thus unreliable for estimating the convergence, and can produce the appearance of individual differences, even if none exist

Study 2: Questions about lexical frequency effects

- Can previous results finding an effect of lexical frequency on convergence be the result of something other than lexically-specific convergence?
- Addressed in a production-only study:
  - The first mention of words is prone to hyperarticulation (Bard et al. 2000; Fowler and Housum 1987); this may be stronger for lower frequency words (Wright 1979), making measured baselines less reliable for them
  - Repetitions of these words are likely to be more natural than the first productions
  - Natural productions will be more similar across speakers, so increased naturalness could create the appearance of convergence when the shift is compared to another speaker as a reference value

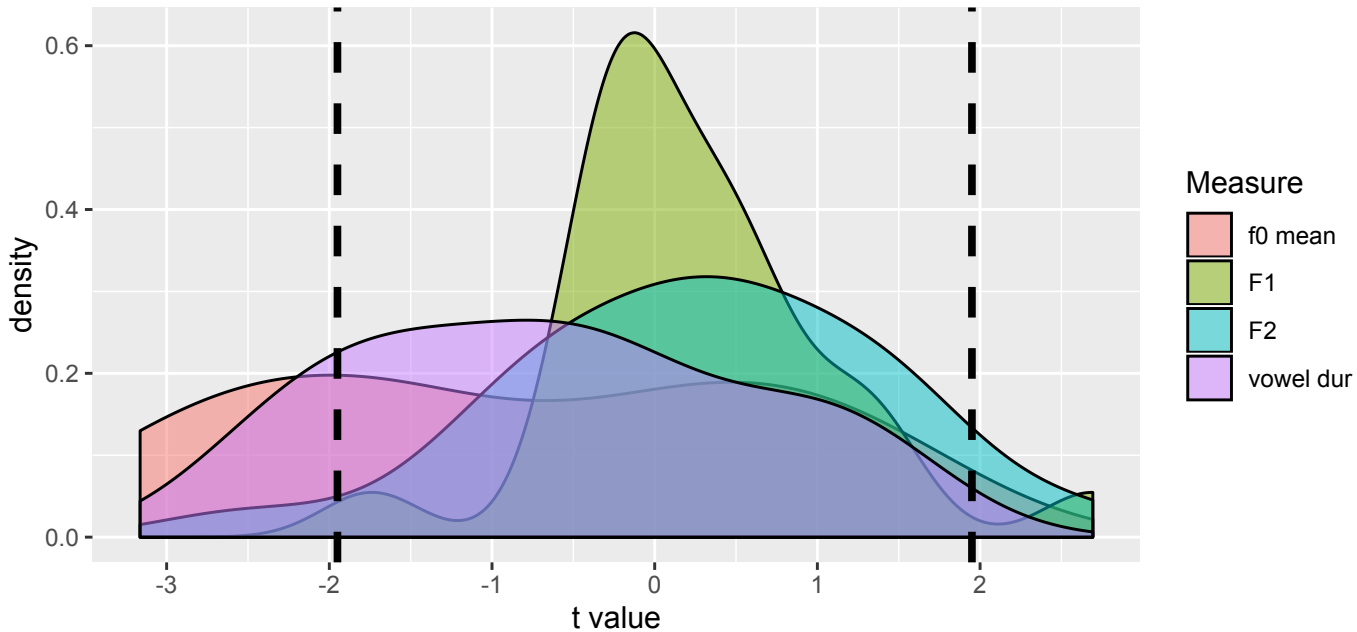
Methods

- **Participants**: 24 female native speakers of American English
- **Task**: Reading a set of 120 English words in randomized order; the full set was produced twice (Note that this is *not* a convergence task – participants were not exposed to any other speaker)
- Words were selected to have an approximately normal distribution of log frequencies
- **Measurements**: For each word, F1, F2, vowel duration, and f0 mean were measured

Methods

- The recordings were treated as if they came from a shadowing experiment, with the first production before exposure to the model talker and the second production after exposure
- There were 24 iterations of each analysis, using each of the participants as the reference value (as if that participant had been a model talker heard by the others)
- Following previous work on frequency effects, "convergence" was measured as change in distance from the reference value:
  $|\ ParticipantStart - Reference\ |$ - $|\ ParticipantFinal - Reference\ |$

Results: Lexical frequency as a predictor of "convergence"



Using initial productions as reference values

Lexical frequency was a significant negative predictor (more convergence with lower frequency words) in 12/96 models and a significant positive predictor in 2/96 models

Study 2 Results Summary

- The results show apparent frequency-conditioned convergence: More convergence with lower frequency
- Mostly in F0 and vowel duration; formant patterns may be complicated by opposing frequency-conditioned reduction and by differences across vowel qualities
- In a word list selected for extreme frequencies, and with more repetitions, the artificial appearance of frequency-based convergence would likely be even more apparent

Study 3: Questions about word-specific convergence

- Given the doubt cast on frequency-dependent convergence effects, is there any evidence for lexically-specific convergence?
- The characteristics investigated are usually consistent across the stimuli, either with the same manipulation (e.g. Nielsen 2011), or naturally produced by the same speaker (e.g. Pardo et al. 2017)
- If word frequency effects reflect the accumulation of exemplars of each word, it should be possible to elicit a shift of the same characteristic in different directions in different words

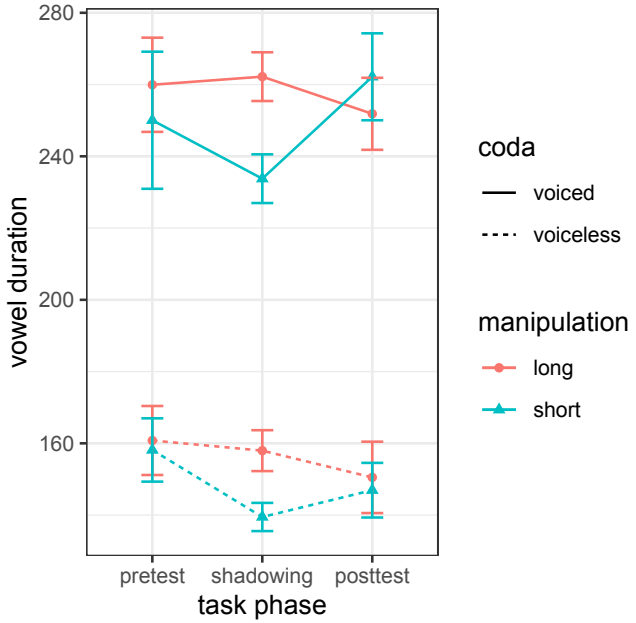
Methods

- **Participants**: 24 female native speakers of American English
- **Task**:
  1. Pretest – Reading a set of 36 English words in randomized order, along with 84 filler words
  2. Shadowing – participants repeated after 36 acoustically manipulated target words, in randomized order; each was presented three times
  3. Posttest – Reading the same original set of words again
- For each characteristic, a listener heard an equal number of words with each manipulation (vowel duration or F2); the manipulation was always the same for the three repetitions of the same lexical item, e.g. *boot, brute, hoot, moose, shoes, zoo* with raised F2, and *boost, cooed, choose, do, fruit, hoop* with lowered F2
- **Measurements**: F2 and vowel duration were measured in each word in each phase of the experiment

Methods

- Reported statistics come from mixed effects linear regression models with each characteristic as the dependent variable
- Pre-task productions were used to establish the speakers' baselines; there was no initial difference between the words used in each condition

Results: Vowel duration manipulations

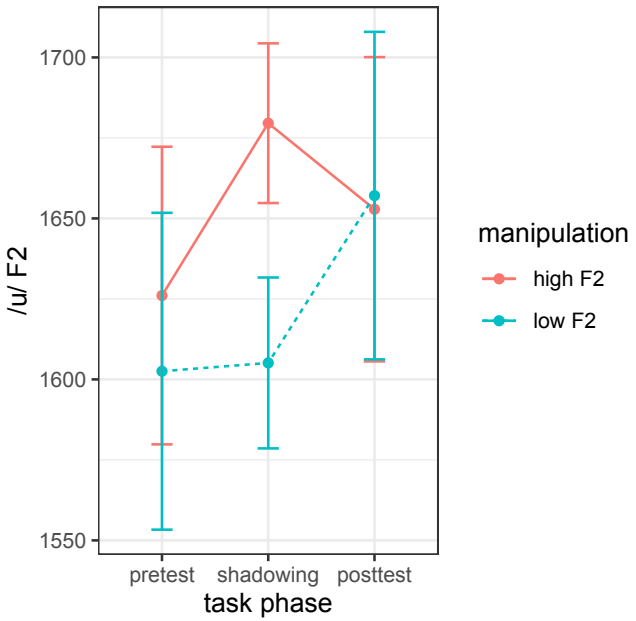


Mean and 95% CI by task phase and manipulation condition for vowel duration manipulations.

Vowel duration in shadowing was longer when repeating after words with lengthened vowels than words with shortened vowels, both with a voiced coda (262 ms vs. 234 ms, $p < 0.0001$) and voiceless coda (158 ms vs. 139 ms, $p < 0.0001$)

But despite the effects in shadowing, there was no effect of manipulation on post-task productions

Results: /u/ F2 manipulations



Mean and 95% CI by task phase and manipulation condition for /u/ F2 manipulations.

In shadowing, F2 in /u/ was significantly higher when repeating after a /u/ word with raised F2 than a /u/ word with lowered F2 (1680 Hz vs. 1606 Hz, $p < 0.0001$)

But despite the effects in shadowing, there was no effect of manipulation on post-task productions

Study 3 Results Summary

- No evidence for lexically-specific effects
- It is possible that this amount of exposure is to little to elicit lexically-specific effects; however, such a limitation also suggests that previous findings of frequency effects have an alternative explanation

Study 4: Questions about generalization across sounds

At what level of the representation does convergence generalize?

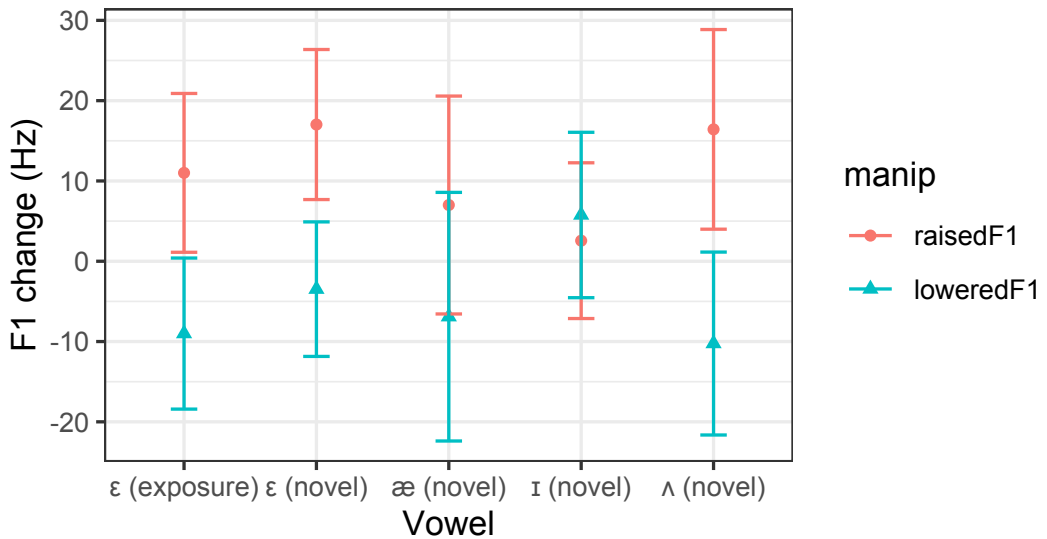
- Nielsen (2011) demonstrated that convergence in VOT can generalize across sounds with the same feature
- However, it is unclear whether or not feature-level effects would behave similarly for other features
- With vowels, would a shared shift be among vowels with a shared target, or would the whole vowel space shift to preserve contrasts?

This study uses a shadowing task to test how exposure to shifted F1 in a single vowel quality (/ɛ/) influences F1 of other vowels that match either in frontness or in height

Methods

- **Participants**: 24 female native speakers of American English
- **Task**:
  1. Pretest – Reading a set of 60 English words in randomized order, along with 60 filler words
  2. Shadowing – participants repeated after 15 acoustically manipulated target words given in randomized order; each was presented three times. The exposure items in this task only had the vowel /ɛ/: *best*, *bet*, *dead*, *debt*, *fed*, *guess*, *less*, *mess*, *met*, *net*, *pet*, *red*, *set*, *test*, *wet*
  3. Posttest – Reading the same original set of words again
- There were two conditions: half of participants heard these words with a raised F1 in the /ɛ/ and half heard a lowered F1
- **Measurements**: F1 in the 15 words from the shadowing task with /ɛ/, in 15 other words with /ɛ/, and in 10 items each with /æ/, /ɪ/, and /ʌ/

Results: F1 change



Participants' F1 increased in the raised F1 exposure condition and decreased in the lowered F1 exposure condition

This effect was present both for words with /ɛ/ (β=20.3, SE=7.47, t=2.71, p=0.0104), and also for words with /ʌ/ (not significantly different from /ɛ/: β=6.41, SE=9.53, t=0.67, p=0.50)

Study 4 Results Summary

- Exposure to F1 manipulation in /ɛ/ is extended to other mid vowel tested in the experiment, /ʌ/
- Suggests that convergence to one vowel is generalized to vowels that have the same target in the domain of manipulation
- Does not produce corresponding shifts in other vowel heights, which start out with different F1 targets

Conclusions: Individual people

- By-interlocutor tendencies: A social effect, cf. previous work demonstrating greater convergence when the interlocutor is viewed positively
- Lack of consistent by-speaker tendencies: There aren't some people who are consistently more convergent than others (no broad cognitive differences driving convergence), though there is variation in which characteristics individual speakers most converge in

Conclusions: Generalization across words and sounds

- No evidence for word-specific convergence
- Convergence generalizes across words with the same sound
- A shift in one sound also influences other sounds with shared targets for the shifted characteristic

Conclusions: Artifacts of measurements

There are artifacts of how convergence is measured: Unreliable baselines can produce effects that are not capturing convergence, and the DID method produces further artifacts:

- Apparent by-speaker individual differences in convergence, correlated with baseline distance from the interlocutor and baseline distance from the population mean
- Apparent by-word differences in convergence; correlation with starting distance results in a correlation with lexical frequency

References

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J Phon, 40*, 177–189.

Babel, M., & Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Lang Speech, 55(2)*, 231–248.

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.

Bane, Max, Peter Graff, & Morgan Sonderegger. (2010). Longitudinal phonetic variation in a closed system. In R. Baglini, T. Grinsell, J. Keane, A.R. Singerman, and J. Thomas (Eds.), *Proc. of the 46th Meeting of the Chicago Ling Society* (pp. 43–58).

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1–22.

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.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory & Language, 49*, 396–413.

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

Hwang, H., & Chun, E. (2018). Influence of social perception and social monitoring on structural priming. *Cognitive Science, 42*, 303–313.

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

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*, 132–142.

Pardo, J.S, Jordan, K., Mallari, R., Scanlon, C., & Eva Lewandowski. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language, 69*, 183–195.

Pardo, J.S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics, 79*, 637–659.

Pardo, J.S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics, 69*, 1–11.

Sanker, C. (2015). Comparison of phonetic convergence in multiple measures. In *Cornell Working Papers in Phonetics and Phonology 2015*, pages 60–75.

Tamminga, M., Wade, L., and Lai, W. (2018). Stability and variability in phonetic flexibility. Talk presented at the Linguistic Society of America annual meeting.

Weise, A., & Levitan, R. (2018). Looking for structure in lexical and acoustic-prosodic entrainment behaviors. In *Proceedings of the 2018 conference of NAACL: Human language technologies* (Vol. 2, pp. 297–302).

Yu, A., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PloS ONE, 8(9)*, e74746.