Phonetic Convergence in Mandarin

Haoru Zhang and Chelsea Sanker
Brown University

Introduction: Convergence is the phenomenon in which people match others' behaviors in a communicative setting. Although a wide range of phonetic *variables* have been identified that exhibit convergence, e.g. F0 (Babel & Bulatov, 2012), formants, and vowel duration (Sonderegger, Bane & Graff, 2017), most previous convergence studies have examined a small number of *languages*, with the vast majority using English.

Some work suggests that phonological contrastiveness limits converge (Podlipský & Šimáčková, 2015) or, conversely, increases convergence based on increasing attention to particular details (Mitterer & Ernestus, 2008). However, comparisons are often within a single language, so differences between measures may be due to the phonological system or may reflect phonologically independent characteristics of each measure. Data from tonal languages, in comparison to previous work on non-tonal languages, can provide insight into the role of phonology in phonetic convergence. The current study expands on previous work by examining controlled experimental data from a widely spoken tonal language—Mandarin.

Methods: 42 Mandarin speakers (age 18-30, 19 female) were recruited in Beijing, China and assigned in two experimental paradigm groups (shadowing and exposure). Participants heard 58 monosyllabic CV Mandarin words spoken by a native Mandarin speaker, split into three blocks. For the shadowing group, participants repeated after each word; for the exposure group, participants listened to a block of words and then read the same words. All participants read the target words aloud before the listening task, produced them three times during the task, and read them again afterwards. F0 maximum, F1, F2, and vowel duration were measured, for all four tones and seven vowels. Data was modeled with linear mixed-effects regression.

Results: The results have been pooled across experimental groups, as there was no effect of task type. Convergence was found in Mandarin similar to that previously observed in other languages; the model talker's productions were predictive of participants' post-task speech. In addition, phonologically contrastive categories shaped convergence effects for F0 maximum, F1 and F2; these measures exhibited both overall convergence and category-specific convergence.

Speakers exhibited significant convergence in F0 maximum (shown in Figure 1), comparable to F0 convergence in English, but additional convergence effects were sensitive to particular tone categories. Speakers also exhibited significant convergence in vowel formants, sensitive to vowel category; F1 is shown in Figure 2 (F2 exhibited similar results, omitted due to space). The main effect suggests convergence shifting the overall vowel space, while vowel-specific effects indicate that category membership shapes convergence. These results parallel previous results in English. Finally, there was also significant convergence in vowel duration, as shown in Table 3. In comparison to formants and F1, it was not sensitive to phonological categories, either tone categories, as illustrated in Figure 3, or vowel categories.

Conclusions: The results offer support for phonetic variation shaped by phonological contrasts, as phonological categories form salient domains within which convergence occurs. There are also broad patterns of convergence within a phonetic characteristic, but convergence can vary within the specific categories that contrast in that measure. While Mandarin speakers, like English speakers, converge broadly to speaking higher or lower in their F0 range, this shift is not reflected equally across tone categories, which may suggest different salience of F0 by category, producing different degrees of attention and subsequent convergence.

	Estimate	Std. Error	df	t value	p value
(Intercept)	2.99	0.32	79.3	9.47	< 0.0001***
Pretest	0.24	0.023	1783	10.97	< 0.0001***
Model Talker (Convergence)	0.17	0.055	64.8	3.19	0.0022**
Tone - Set 1	0.79	0.34	57.7	2.32	0.024*
Tone - Set 2	-0.92	0.63	57.2	-1.47	0.15
Tone - Set 3	-0.06	0.37	69.5	-0.17	0.86
Model * Tone Set 1	-0.14	0.062	57.7	-2.29	0.026*
Model * Tone Set 2	0.16	0.12	57.3	1.42	0.16
Model * Tone Set 3	0.002	0.067	69.4	0.03	0.98

Figure 1. Regression model for post-task F0-maximum. Intercept : Tone = Grand mean of tones.

	Estimate	Std. Error	df	t value	p value
(Intercept)	0.038	0.039	376.7	0.97	0.33
Pretest	0.25	0.026	1833	9.8	< 0.0001***
Model Talker (Convergence)	0.45	0.097	34	4.7	< 0.0001***
Vowel - Set 1	0.88	0.15	16.6	5.7	< 0.0001***
Vowel - Set 2	0.048	0.049	169.7	0.97	0.33
Vowel - Set 3	-0.37	0.084	22.1	-4.45	< 0.0001***
Vowel - Set 4	-0.015	0.079	122.3	-0.19	0.85
Vowel - Set 5	0.37	0.15	563.6	2.41	0.016*
Vowel - Set 6	-0.29	0.061	138.9	-4.78	< 0.0001***
Model * Vowel Set 1	-0.34	0.10	733.2	-3.25	0.0012**
Model * Vowel Set 2	-0.11	0.11	547.8	-0.95	0.34
Model * Vowel Set 3	-0.11	0.10	1029	-1.12	0.26
Model * Vowel Set 4	-0.04	0.25	954.3	-0.16	0.87
Model * Vowel Set 5	1.35	0.41	1330	3.31	< 0.0001***
Model * Vowel Set 6	-0.33	0.12	189.2	-2.81	0.0054**

Figure 2. Regression model for post-task F1. Intercept : Vowel = Grand mean of vowels.

	Estimate	Std. Error	df	t value	p value
(Intercept)	0.13	0.026	57.2	4.93	< 0.0001***
Pretest	0.22	0.025	1623	8.98	< 0.0001***
Model Talker (Convergence)	0.35	0.088	50.7	3.97	< 0.0001***
Tone - Set 1	-0.046	0.055	44.7	-0.84	0.4
Tone - Set 2	-0.032	0.046	44.7	-0.69	0.49
Tone - Set 3	0.06	0.031	47.9	1.97	0.055 .
Model * Tone Set 1	0.13	0.185	44.3	0.7	0.49
Model * Tone Set 2	0.24	0.167	44.3	1.44	0.16
Model * Tone Set 3	-0.13	0.094	44.5	-1.33	0.19

Figure 3. Regression model for post-task vowel duration. Intercept : Tone = Grand mean of tones.

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech, 55*(2), 231-248.

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168-173.

Podlipský, V. J., & Šimáčková, Š. (2015). Phonetic imitation is not conditioned by preservation of phonological contrast but by perceptual salience. *International Congress of Phonetic Sciences (ICPhS) 2015*.

Sonderegger, M. Bane, M. & Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language, 93(3)*, 598-640.