

Phonetic convergence of Hong Kong English: sound salience and the exemplar-based account

Wenling Cao (Grace)

University of York, UK
gracecao325@gmail.com

ABSTRACT

Previous theories of convergence either focus on the motivation of convergence or address the autonomy of convergence in a dialogue. However, neither of them provides a clear mechanism of how convergence occurs. The present study tries to answer the question by proposing the exemplar-based account of speech convergence. It argues that a speaker's distributions of sound properties keep updating when he/she receives new input from the environment, and the more salient a sound is, the more it accommodates. An empirical study of Hong Kong English (HKE) was conducted to examine the account. Nineteen HKE speakers completed a map task with an RP speaker and with a General American English speaker separately for one hour. Their production of THOUGHT and BATH vowels, rhoticity, fricative /z/ and fricative /θ/ were examined before, during and after the map tasks. Significant convergence was found on rhoticity and fricative /z/, which supported the predictions of the exemplar-based account.

Keywords: short-term accommodation, exemplar theories, perception and production

1. INTRODUCTION

Communication Accommodation Theory [1] (henceforth CAT) has been a prevalent theory in speech accommodation since the 1990s. It suggests that people accommodate to each other in order to shorten the social distance between them. As a theory developed from social psychology, CAT tends to explain why people accommodate instead of how people accommodate. Pickering and Garrod [2] proposed the interactive-alignment model, suggesting that convergence is an automatic process in a dialogue. Though the interactive-alignment model indicated that alignment is achieved through “a primitive and resource-free priming mechanism”, it is still unclear that how exactly the priming mechanism operates.

It seems that neither CAT nor the interactive-alignment model could answer the question: how does convergence occur? The present study adopts the exemplar-based account of speech convergence,

aiming to provide some preliminary thoughts on the question. A convergence study on Hong Kong English (HKE) is reported too.

2. THE EXEMPLAR-BASED APPROACH

The exemplar-based approach began in the 19th century in psychology of memory and has been adopted in many fields; for example, Johnson [3, 4] in speech perception and Pierrehumbert [5, 6] in speech production.

2.1. Exemplar models in speech perception

Johnson [3] suggests that categorisation is achieved by comparing a new input item with each of the remembered instances/exemplars of each category. The similarity between the item and each exemplar determines the activation level of the exemplar. The greater the similarity, the higher the activation level. The new item should be categorised as an example of the category which has the highest activation level after comparing it to other categories. When people perceive speech from other speakers, social categories (e.g. a speaker's identity, gender and accent) associated with specific phonetic properties are also stored as part of a general learning process.

2.2. Exemplar models in speech production

Pierrehumbert [5, 6] proposes an exemplar model of speech production, trying to model a complete perception-production loop based on exemplar representation. In production, a cognitive concept (e.g. I want to say “bath”) is created first, then the signal passes down to production system to select the relevant label. A random sampling of the exemplar distribution is taken for the label. With the neighbourhood region of the selected exemplar activated, the average properties of the region form the production goal.

Pierrehumbert [5, 6] extends Johnson's exemplar-based perception model to production. Her model consists of a complete perception-production loop. This is an important step for studies like speech accommodation, which makes it possible to explain how perceptual input from the interlocutor affects a speaker's pronunciation.

2.3. Exemplar-based account of speech convergence

In an ideal situation, the exemplar-based account of speech convergence proposes that a speaker starts with his/her own property distributions of each exemplar that he/she has developed based on the previous experience. When the speaker talks to an interlocutor, the distributions keep updating based on the input. As the production goal is selected and calculated from the updated distributions, the speaker's pronunciation should shift towards the interlocutor's. In this way, the speaker converges towards the interlocutor. Although the classic study of Goldinger [7] was designed to examine the exemplar theory in speech imitation, Goldinger's study focused on the frequency effect and the role of social factors was not considered.

In a real conversation, a speaker faces more challenges than merely updating the distributions of his properties. Social factors also play a role in the updating process. For example, if a speaker associates certain sounds with his regional identity, the updates of these sounds might be inhibited when he wants to signal his regional identity in the conversation.

3. A CASE OF HONG KONG ENGLISH

A study of HKE is reported here and the exemplar-based account is used to explain the results. 19 HKE speakers (12 females) who were studying in the UK attended experiments. Their English proficiency was range from IELTS 6.0 to 8.0. Two female RP speakers and two female General American English (henceforth GenAmE) speakers were recruited as native interlocutors.

3.1. Experiment design

The HKE speakers completed a pre-task, a map task and a post-task in a recording booth. In the pre-task and the post-task, the participants were given three maps and were asked to describe what they can see on the maps. They completed these two tasks by themselves in the recording booth. In the map task, the HKE participants talked to a native speaker (either with an RP speaker or with a GenAmE speaker). By following the native interlocutor's instructions, they drew a route and corrected the wrong landmarks on the maps. 10 participants talked to an RP speaker first and then repeated the map task again with a GenAmE speaker; the rest of the participants completed the map tasks with the native speakers in a reverse order.

3.2. Variables

The variables for the study were the THOUGHT vowel, BATH vowel, rhoticity, /z/ and /θ/. Table 1

shows pronunciations of the five variables in HKE, RP and GenAmE respectively. The vowels were chosen because they are pronounced differently in RP and GenAmE. Rhoticity was chosen because HKE does not have a consistent pattern of rhoticity. Some HKE speakers carry rhoticity in their production while some do not. The fricatives were selected because they are typical features of HKE, where fricatives /z/ and /θ/ are usually pronounced as [s] and [f] in HKE.

Based on the exemplar-based account, the HKE participants were expected to converge towards the accent they were exposed to in the map tasks.

Table 1: Five variables of the study.

Variable	HKE	RP	Gen AmE	Keywords (e.g.)
THOUGHT	[ɔ]	[ɔ]	[ɑ]	<i>cause, small</i>
BATH	[a]	[ɑ:]	[æ]	<i>pass, bath</i>
rhoticity	yes/no	no	yes	<i>car, star</i>
/z/	[s]	[z]	[z]	<i>zoo, zero</i>
/θ/	[f]	[θ]	[θ]	<i>three, thirty</i>

3.3. Data analysis

In total 14,009 target words containing THOUGHT and BATH were annotated in Praat [8] for both the HKE participants and the native interlocutors in the three tasks. F1 and F2 values were extracted from the midpoint of the annotated vowels using Praat scripts. Labov's method [9] was used for vowel normalisation. For rhoticity, 9,568 words that contain [ɹ] sounds in syllable-coda position were annotated. For fricatives, 7,084 words containing /z/ and 6,730 words containing /θ/ in word-initial position were annotated. An auditory judgement was made on each annotated consonant by the author, and checked by another trained phonetician from the University of York. The agreement tests showed an 89% of agreement for rhoticity, 93% for fricative /z/ and 78% for fricative /θ/.

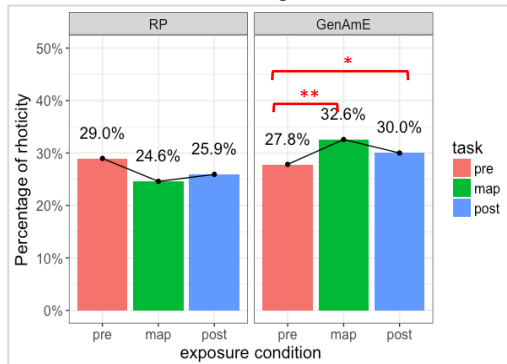
3.4. Result

3.4.1. Consonants

For consonantal variables, percentages of rhoticity, fricative [z] and fricative [θ] were calculated in the three tasks. Figure 1 shows the results for rhoticity. From the pre-task to the map task, the HKE speakers produced fewer rhotic words in the RP condition (decreased by about 4%), whereas in the GenAmE condition more rhotic words were produced (increased by about 5%). Logistic mixed effects regressions were run using a binary judgement of rhoticity as the dependent variable. The full model

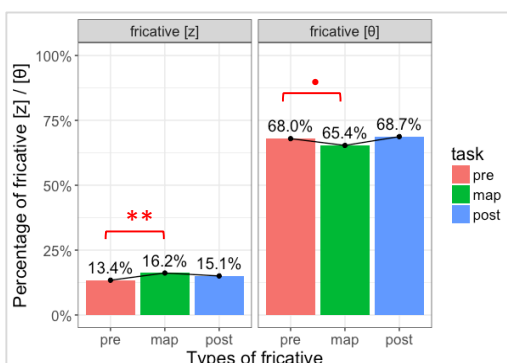
included task (*pre*, *map*, *post*), exposure (*RP*, *GenAmE*), the interaction between task and exposure, and participant sex as fixed effects. It also included random intercepts by participant, by word (*referred to the words that contained the target vowels*) and by interlocutor, and random slopes by participant over task, exposure and the interaction between task and exposure. A nested model with task*exposure removed was compared with the full model using ANOVA. The results suggested that **task*exposure** was a significant predictor ($\chi^2 = 9.79$, $DF = 2$, $p = .007$), indicating that the HKE participants changed their rhoticity depending on the accents they were exposed to.

Figure 1: Percentages of rhoticity across the three tasks and two conditions. * indicates significant comparisons based on post-hoc tests of regressions at 0.05 level; ** indicates a significance at 0.01 level.



For the two fricatives, logistic mixed effects regressions were run. In the full model of fricatives, exposure was excluded from the fixed effects because RP and GenAmE have the same realisation for fricative /z/ and fricative /θ/. Therefore, the full model contained task and sex as the fixed effects, it also included random intercepts by participant, by word and by interlocutor, and random slope by participant. The results suggested a significant convergence on [z] and a marginally significant divergence on [θ] from the pre-task to the map task, as shown in Figure 2.

Figure 2: Percentages of fricatives across the three tasks and two conditions. *** indicates a significant comparison based on the post-hoc tests of regressions at 0.001 level; · indicates a marginal significance.



3.4.2. Vowels

For the THOUGHT and the BATH vowels, Euclidean distances between the HKE participants and the native interlocutors in the three tasks were calculated. Linear mixed effects regressions were run using Euclidean distance as the dependent variable for each vowel. The fixed effects, random effects and random slopes in the full model of the vowels are the same for those in the full model of rhoticity. A nested model with task*exposure removed was compared with the full model.

For the THOUGHT vowel, the interaction of **task*exposure** significantly improved the model fit, indicating that the participants' changes in Euclidean distance from the pre-tasks to the map tasks in the RP condition were significantly different from the changes in the GenAmE condition. Post-hoc tests suggested that no significant changes from the pre-task to the map task/post-task were found in both conditions.

For BATH vowel, task*exposure was not significant, however, **task** was a significant predictor in the model ($\chi^2 = 19.028$, $DF = 2$, $p < .001$). Post-hoc tests suggested that significant divergence was found from the pre-task to the map task in both conditions.

4. DISCUSSION

Although it is difficult to summarise a unique pattern from the complex results above, there is some evidence suggesting a correlation between the salience of sounds and accommodation.

4.1. Salience and accommodation

Salience is defined as sounds which have a greater phonetic difference between the HKE speaker's native repertoire and the native interlocutor's repertoire, and sounds which carry specific social meanings in the present study.

In the current study, convergence was found on rhoticity and fricative [z], but not on fricative [θ]. This might be due to rhoticity and fricative [z] having a greater phonetic difference between HKE and RP/GenAmE than fricative [θ]. Firstly, rhoticity is a more salient feature compared to fricative [z] and fricative [θ] for HKE speakers, because convergence on rhoticity involves adding or deleting a phoneme /-r/, while convergence on the fricatives only involves replacement.

Secondly, fricative [z] is more salient than fricative [θ] for HKE speakers because the phonetic differences between /z/-/s/ are larger than the differences between /θ/-/f/. Jongman, Wayland and Wong [10] found that /s/ and /z/ were significantly different in duration, amplitude and spectral peaks,

whereas /f/ and /θ/ were similar in duration and spectral peaks, and the main difference between /f/ and /θ/ was on F2 transition information only. The HKE participants might be more easily to perceive fricative [z] than fricative [θ] because the /s/-/z/ differences are larger than the /f/-/θ/ differences. Babel [11,12] also suggests that people tend to converge on the vowels which have a larger acoustic-phonetic distance from the model talkers.

Salience can also refer to sounds which carry social meanings, for example, rhoticity. The HKE speakers were found to converge towards the RP/GenAmE speakers on rhoticity in the map tasks. This might be also due to rhoticity/non-rhoticity being a stereotype for American English/British English. Instead of recognising that the [æ] vowel in the word “bath” is a feature of GenAmE, it might be easier for HKE speakers to associate rhoticity with American English.

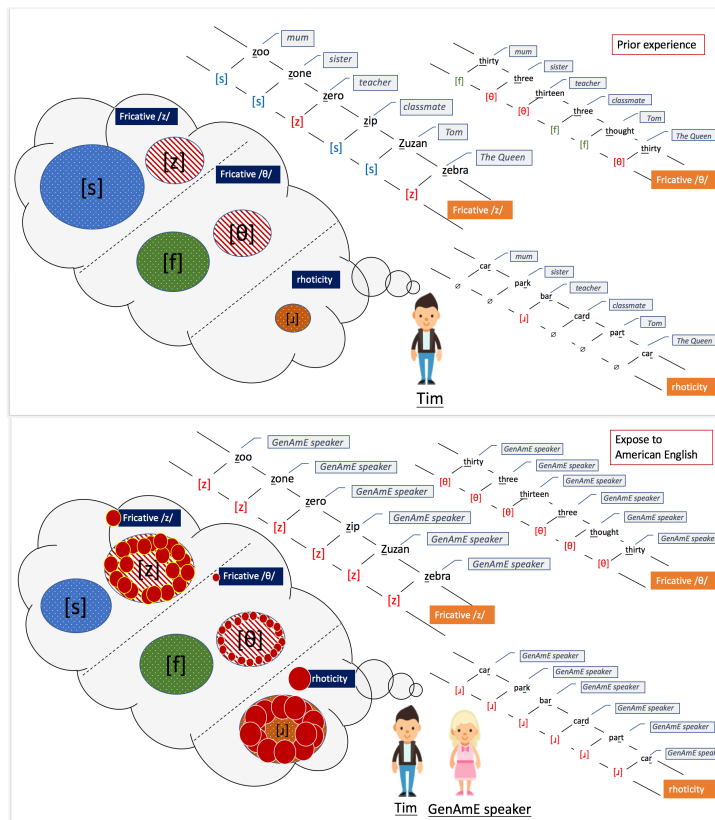
4.2. The exemplar theories

If we adopt the exemplar-based account to explain the accommodation of the consonants, three examples are given in Figure 3. The HKE speaker in Figure 3 is called *Tim*. Presumably before Tim arrives the UK, he has established different distributions for fricative /z/, fricative /θ/ and rhoticity after living in Hong Kong for 24 years. For example, Tim has established a category of voiceless [s] representing the HKE variant, and a category of voiced [z] representing the standard variant for the phoneme /z/. The [s]-category would be more dominant at this point because he would receive more input of [s].

When Tim talks to a GenAmE interlocutor in the map task, his distribution of exemplars containing [z] keeps updating and the probability of [z] being selected for production increases because he is receiving input of [z] from the GenAmE speaker. According to the previous analysis of the salience effect (degree of salience: rhoticity > fricative [z] > fricative [θ]), the input of rhoticity would be weighted greater than the input of fricatives [z] and [θ] for Tim, which results in a larger change for rhoticity than for the two fricatives. In addition, an analysis of the input showed that the HKE participants on average received more native tokens of rhoticity (mean = 266 tokens) and fricative [z] (mean = 187 tokens) from the native interlocutors than the other sounds (mean < 140 tokens for the others). In other words, not only the quality (e.g. weight) but also the quantity of the rhotic input contribute to the convergence.

Moreover, the associated social-index label (e.g. *rhoticity* – American English) would increase the probability of rhoticity being selected as it also contributes to a greater weight in the input.

Figure 3: Illustration of the exemplar-based account for convergence. The dotted and stripped circles represent the categories of the consonants. For example, the circles of [s] and [z] together represent the distribution of the z-sound. The larger the circle, the more likely the sound in the circle is to be selected as the production goal. Each token of input is represented by a red dot. The larger the dot is, the greater weight it has during the update of distribution.



What the account could not explain is the divergence of fricative /θ/. According to the account, the HKE speakers should also show convergence in fricative /θ/, yet, the result indicated a marginal divergence. One possibility is that the HKE speakers might perceive the fricative /θ/ as [f] due to the perceptual bias affected by their L1-Cantonese, which might instead increase the probability of [f] being selected. This reveals one of the limits of the account, that is, speakers do not necessarily perceive the input as identical to what their interlocutors pronounce.

5. CONCLUSIONS

The present study reported a phonetic convergence study of HKE towards RP/GenAmE and the exemplar-based account was used to explain some of the results. Supporting evidence was found on the HKE participants' convergence on rhoticity and fricative [z].

6. REFERENCES

- [1] Giles, H., Coupland, J., Coupland, N. 1991. *Contexts of Accommodation: Studies in Emotion & Social Interaction*. Cambridge: Cambridge University Press.
- [2] Pickering, M. J., Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, 27(2), 169-190.
- [3] Johnson, K. 1997. Speech perception without speaker normalization: an exemplar model. In: Johnson, K., Mullennix, J.W. (eds), *Talker Variability in Speech Processing*. San Diego: Academic Press, 145-166.
- [4] Johnson, K. 2007. Decisions and mechanisms in exemplar-based phonology. In: Sole, M, J., Beddor, P., Ohala, M. (eds), *Experimental Approaches to Phonology. In Honor of John Ohala*. Oxford: Oxford University Press, 25-40.
- [5] Pierrehumbert, J. B. 2001. Exemplar dynamics: word frequency, lenition and contrast. In: Bybee, J., Hopper, P. (eds.), *Frequency Effects and the Emergence of Linguistic Structure*. Amsterdam: John and Benjamins, 137-158.
- [6] Pierrehumbert, J. B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46(2-3), 115-154.
- [7] Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251-279.
- [8] Boersma, Paul., Weenink, David. 2017. Praat: doing phonetics by computer [Computer program]. version 6.0.28, retrieved 6 August 2017 from <http://www.praat.org/>.
- [9] Labov, W., Ash, S., Boberg, C. 2006. *The Atlas of North American English*. Berlin: Mouton de Gruyter.
- [10] Jongman, A., Wayland, R., Wong, S. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3), 1252-1263.
- [11] Babel, M. E. 2010. Dialect divergence and convergence in New Zealand. *Language in Society*, 39(4), 437-456.
- [12] Babel, M. E. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177-189.