

MUTUAL INTELLIGIBILITY OF ENGLISH VOWELS BY CHINESE DIALECT SPEAKERS

Rongjia Cui & Vincent J. van Heuven

Phonetics Laboratory, Leiden University Centre for Linguistics, the Netherlands

r.cui@hum.leidenuniv.nl; v.j.j.p.van.heuven@hum.leidenuniv.nl

ABSTRACT

This mutual intelligibility study contains a production and perception experiment on English vowels by Chinese learners. In the production experiment, 45 male first-year Chinese college students were recorded. They hailed from nine different dialectal backgrounds (three supergroups), with five speakers per dialect group. The stimuli were [hi:d], [hɪd], [hed], [hæd], [hu:d], [hʊd], [hɔ:d], and [hɒd]. Formants as well as durations were measured. Linear Discriminant Analyses showed that the speakers' dialect backgrounds can be predicted better than chance, but only at the supergroup level. In the perception part, one representative male speaker was chosen for each dialect based on his Euclidean distance from a model American speaker. The representative vowel tokens were then identified and rated for typicality by 282 first-year undergraduates from the same nine dialect groups. A significant interlanguage benefit (i.e. better identification results when listener and speaker share the same language background) was found, but again only on the dialect supergroup level.

Keywords: mutual intelligibility, interlanguage speech intelligibility benefit, Chinese dialects, English vowels

1. INTRODUCTION

It is widely recognized that a speaker's native language (L1) or dialect influences his/her second-language (L2) production. The L1 transfer can be either negative or positive. The current research not only addresses the L1 influence (in this study: Chinese) on L2 (i.e. English) production, but also on L2 perception. In particular, in a production experiment we aim to find out to what extent a speaker's native dialect interferes with his/her English pronunciation, i.e. is it possible to determine a Chinese speaker's dialectal background from his/her pronunciation of the English vowels? In a perception experiment, we ask what

is the mutual intelligibility of English vowels by Chinese dialect speakers, i.e. how well do Chinese listeners identify English vowels produced by Chinese EFL (English as a Foreign Language) speakers with same or different native dialect background.

The idea behind these experiments is that even though Chinese dialects have many characteristic pronunciation features in common, and sound very much the same to Western listeners, they are linguistically as different from one another as certain European languages within the same phylum (such as Spanish, Portuguese and Italian). To the extent that the native L1 phonology influences the pronunciation of the L2, we would predict that Chinese learners of English should be able to discern, just by listening, whether another Chinese learner of English hails from the same or a different dialect background. Also, it should be possible to find specific acoustic characteristics in the varieties of English spoken by Chinese learners with different native dialect backgrounds, that would allow us to pinpoint the native dialects.

Moreover, when learners of a foreign language who share the same native language (or dialect in our case), communicate with each other in the foreign language, their mutual intelligibility should be better than when speaker and listener do not share the same language background. This prediction was confirmed in earlier research with foreign learners of English from different language families, such as Chinese, Korean and miscellaneous other languages [2], or Dutch and Chinese [4, 8, 10]. However, this so-called interlanguage speech intelligibility benefit (hence ISIB) has not yet been shown to exist between learners of a foreign language who share the same native dialect (or whose native languages are closely related varieties within a family). It is the purpose of the present study to determine to what extent an ISIB effect may also be found at the dialect level within a language.

2. PRODUCTION EXPERIMENT

In this experiment we recorded 45 male first-year undergraduate students, who produced readings of eight English words, namely, *heed* [hi:d], *hid* [hid], *head* [hed], *had* [hæd], *who'd* [hu:d], *hood* [hud], *hawed* [hɔ:d], and *hod* [hɒd]. These are the eight pure monophthongs of English contained in a /h_d/ consonant frame. These speakers originated from nine different dialectal areas. These are Beijing, Chongqing, Jilin, Shandong, Sichuan, Guizhou, Gansu (all seven dialects belong to the Mandarin (i.e. Northern Sinitic) linguistic supergroup), Jiangxi (a dialect representing the Gan supergroup, i.e. Southern Sinitic), and Fujian (belonging to the Min supergroup, which also belongs to the Southern-Sinitic branch). For a more detailed discussion of Chinese dialect genealogy we refer to [6, 11].

Formants F1 (representing vowel height) and F2 (representing backness and rounding) of these embedded vowels were measured at the temporal midpoint of the target vowel tokens using the Burg algorithm implemented in the Praat speech processing software [3]. Formant tracks were overlaid on spectrograms. Whenever tracks did not follow the formants in the spectrogram, the model-order of the algorithm was changed, until a satisfactory match was obtained. Also vowel duration was measured from the onset of voicing in the vowel to the termination of intensity in the formants. Hertz values were converted to Bark [7] as this corresponds to auditory perception. In order to be able to abstract from speaker-individual differences in formants F1 and F2 were Z-normalized as in [9]. Z-normalisation was shown to be the most satisfactory procedure in the comparison of speakers across dialects [1]. These data were submitted to Linear Discriminant Analysis (LDA [5]), in which we automatically classified the 45 speakers' native dialect background from the F1, F2 and duration values obtained for each of the 8 vowel types, i.e. using an initial predictor set of 24 to discriminate 9 categories.

The results of the LDA showed that the speakers' native dialects could not be discriminated above chance level. In a second attempt, we selected the Beijing group as the representative of all the Mandarin groups since Beijing Mandarin is the closest one to Standard Mandarin and in some situations it is regarded as Standard Mandarin. This reduced the dataset to three dialects, at the

supergroup level, i.e. Mandarin (i.e. Beijing), Gan (i.e. Jiangxi) and Min (i.e. Fujian). The results of this second attempt reveal that speakers' dialectal background can be discriminated above chance level, but only at the level of the dialect supergroup, as shown in table 1.

Table 1: Actual and predicted speaker origin (in terms of linguistic super-group) by LDA (in %).

Actual dialect group (down)	Predicted dialect group		
	Mandarin	Min	Gan
Mandarin	40	0	60
Min	0	100	0
Gan	60	20	20

As seen, the 5 speakers from the Min supergroup (Fujian) form a homogenous group of speakers that differ clearly from the other two dialect supergroups. The 5 Min speakers are correctly classified without a single error. The automatic classification of the other ten speakers contains quite a few errors but only one of these is incorrectly identified as Min. More specifically, 3 out of 5 Gan speakers are classified Mandarin, and 3 out of 5 Mandarin speakers are predicted to be Gan. This shows that Gan and Mandarin are more like each other than either of these groups is similar to Min.

We conclude that a Chinese speaker's dialectal background can be predicted from acoustic properties of his/her EFL (English as a Foreign Language) vowels, but only at the supergroup level

3. PERCEPTION EXPERIMENT

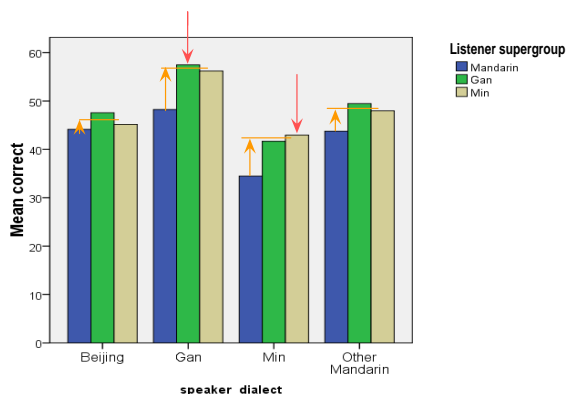
In the production experiment, the Euclidean distances of 45 males' eight vowels as in the above text from the model Standard American English voice (a 29-year-old educated male, speaking General American) were computed.¹ The distances were computed in the bark-transformed and individually Z-normalized vowel space defined by F1 and F2. Duration was not included in the distance measurement since duration was never found to discriminate between the Chinese learners of English. The speakers within each of the 9 dialect groups whose mean distance (across the eight vowel types) from the model speaker was closed to the group average, was taken as the optimally representative speaker of his dialect group. His readings of all 19 pure vowels and diphthongs of English served as the stimuli for the perception test.² These comprised the 8 monophthongs used in the production experiment as

well as the other eleven vowels (*hawed* [hɑ:d], *hard* [hɑ:(r)d], *heard* [hɜ:(r)d], *hayed* [heɪd], *hoed* [həʊd], *hide* [haɪd], *hoyed* [hɔɪd], *how'd* [haʊd], *here'd* [hɪəd], *hoored* [hʊəd], *haired* [heəd]). All vowels had been recorded during the same session described in section 2.

The listeners were 282 first-year undergraduates from the same 9 dialect groups as those the speakers hailed from. None of the speakers served as listeners. There were 31 Min (Fujian) listeners, 25 Gan (Jiangxi), and the remaining 226 listeners belonged to the Mandarin supergroup. Amongst the latter, 36 were Beijing speakers. The perception experiment was conducted in a lecture rooms with fairly large groups of listeners, with stimuli presented over good quality public address loudspeakers, and the listeners responding using pen and paper. At the beginning of the experiment listeners were familiarized with the 19 response categories on their answer sheets. They were then asked to choose on their answer sheet the single best alternative from the set of 19, for each stimulus vowel they heard, with forced choice.

In the analysis of the results we decided to collapse the data for all non-Beijing speakers of Mandarin. All Mandarin listeners (Beijing and other groups alike) were lumped together as a single listener (super)group. Figure 1 summarizes the results of the perception experiment. It presents percent correctly identified vowels broken down by four speakers groups (Beijing, other Mandarin speakers, Gan, Min), and broken down further by listener group (Mandarin, Gan, Min).

Figure 1: Correctly identified English vowels (%) broken down by speaker and listener groups (see text).



The results shown that, across listener groups, the English vowels of the Min speaker are identified correctly least often, while the Gan speaker's vowels are recognized best. The correct

identification rates for the English vowels produced by Mandarin speakers, whether Beijing or other, is intermediate.

Moreover, Figure 1 reveals that, overall, Gan listeners did better than Min listeners, and that Mandarin listeners did poorest of all. However, in order to quantify the ISIB interlanguage speech intelligibility benefit (ISIB) for specific combinations of speaker and listener groups we must look at the scores in relative terms. This means to investigate whether a Min listener hears a Fujian speaker better than, e.g. a Beijing speaker, when communicating in a foreign language. In order to achieve this, we first predict the correct identification score from the mean effect of speaker group and listener group. The Relative ISIB (RISIB) is then the discrepancy between the score predicted from the main effects and the actual score [4].

As can be seen from the length of the upward orange arrows in the second cluster (Gan speaker) and third cluster (Min speaker), the Gan and Min speaker-listener combinations show an ISIB effect, even in absolute terms. The representative Gan speaker was heard best by Gan listeners and the Min speaker was heard best by the Min listeners.

Also both Gan and Min speakers were heard better by each other than by Beijing and other Mandarin listeners. We call this relative ISIB. One reason could be the geographical adjacency of Gan and Min people. Jiangxi Province is to the west of Fujian Province.

Table 2 presents the computation of the relative ISIB measure.

Table 2: Computation of relative ISIB. Effect of listener group (Mandarin, Gan, Min) and of speaker group (Beijing, other Mandarin dialects, Gan, Min), predicted and observed correct vowel identification scores, and the relative interlanguage speech intelligibility (RISIB) are listed. Cases of predicted positive RISIB are highlighted in grey.

List.	Speaker	+/- effect of		pred.	Obs.	RISIB
		list.	speak.			
Mand.	Beijing	-4	-1	42	44	+2
	Other M.	-4	0	43	44	+1
	Gan	-4	+7	50	48	-2
	Min	-4	-7	36	35	-1
Gan	Beijing	+2	-1	48	48	0
	Other M.	+2	0	49	49	0
	Gan	+2	+7	56	57	+1
	Min	+2	-7	42	42	0
Min	Beijing	+1	-1	47	45	-2
	Other M.	+1	0	48	48	0
	Gan	+1	+7	55	56	+1
	Min	+1	-7	41	43	+2

The table shows small positive RISIB for all combinations of speaker and listener groups that share the background dialect: +2, +1 and +2 points for the Mandarin, Gan and Min shared background groups. Smaller positive RISIB is seen in two cases where speaker and listeners do not speak the same dialect but a dialect that is closely related to that of the other, i.e. two Mandarin dialects or two Southern dialects. The RISIB is zero or negative in all other speaker-listener combinations.

4. CONCLUSION

The results of the present study show that Chinese learners of English pronounce the English vowels differently depending on their native dialect background. The effects are small, and cannot be found at the fine-grained level of the dialect itself but only at the cruder level of the dialect supergroup. This was shown by an acoustic analysis of the English vowels followed by automatic vowel classification by Linear Discriminant Analysis. It was also shown by the results of a perceptual vowel identification task performed by groups of Chinese learners of English.

Although the effects are small, they are systematic. As far as we are aware, ours is the first study to show that the interlanguage speech intelligibility benefit can be found not only for speakers and listeners who share the same native language but even when they share the same dialect. Since the differences between dialects tend to be smaller than between languages, it comes as no surprise that the effects are small and subtle.

5. REFERENCES

- [1] Adank, P. 2003. *Vowel Normalization: A Perceptual-Acoustic Study of Dutch Vowels*. Ph.D. dissertation, Nijmegen University.
- [2] Bent, T., Bradlow, A.R. 2003. The interlanguage speech intelligibility benefit. *J. Acoust. Soc. America* 114, 1600-1610.
- [3] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5(9/10), 341-345.
- [4] van Heuven, V.J., Wang, H. 2007. Quantifying the interlanguage speech intelligibility benefit. In Barry, W., Trouvain, J. (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences Saarbrücken Saarbrücken*: Universität des Saarlandes, 1729-1732.
- [5] Klecka, W.R. 1980. Discriminant analysis. *Sage University Paper Series on Quantitative Applications in the Social Sciences*. Beverly Hills, CA: Sage Publication. 7-19.
- [6] Tang, C., van Heuven, V.J. 2009. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua* 119, 709-732.

- [7] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *J. Acoust. Soc. Am.* 88, 97-100.
- [8] Wang, H. 2007. *English as a Lingua Franca. Mutual Intelligibility of Chinese, Dutch and American English*. LOT dissertation series 147, Utrecht, LOT.
- [9] Wang, H., van Heuven, V.J. 2006. Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers. In van de Weijer, J.M., Los, B. (eds.), *Linguistics in the Netherlands*. Amsterdam/Philadelphia: John Benjamins, 237-248.
- [10] van Wijngaarden, S.J. 2001. Intelligibility of native and non-native Dutch speech. *Speech Comm.* 35, 103-113.
- [11] Yan, M.M. 2001. *Introduction to Chinese Dialectology*. LINCOM Studies in Asian Linguistics. München: LINCOM.

¹ This native speaker was born and raised in Des Moines, lived in Iowa City for three years and moved to Boston at twenty-two. He had arrived in the Netherlands in September 2007, about three months before the recordings were made.

² In the analyses of production test, we did not include three vowels [hɔd], [hɑ:(r)d], and *heard* [hə:(r)d] as many American English speakers do not distinguish [hɔ:d] from [hɔd] and *r*-coloring really affects the realization of [ɑ:] and [ə:].