

FORMANT DYNAMICS OF /ua/ IN THE SPEECH OF MANDARIN-SHANGHAINESE BILINGUAL IDENTICAL TWINS

Donghui Zuo & Peggy Pik Ki Mok

The Chinese University of Hong Kong, Hong Kong

donghui@cuhk.edu.hk; peggymok@cuhk.edu.hk

ABSTRACT

This study analyses the formant dynamics of Mandarin and Shanghainese /ua/ in bilingual identical twins. The results show that despite the same organic structure and similar language environment, differences can be found in their formant dynamics. A comparison between the subjects' dominant and non-dominant language shows that the twins are more similar in their dominant language. The data also suggests that separated twins are not necessarily more different than non-separated twins. Their willingness to be twins can affect the similarity of their voices.

Keywords: speaker identity, formant dynamics, identical twins, bilingual

1. INTRODUCTION

Between-speaker difference can be used as a parameter to discriminate speakers. However, due to the flexibility of speech, in addition to between-speaker differences, within-speaker differences can also be found in natural speech [5, 6]. Therefore, the boundary between these two kinds of differences is one of the central issues in speaker-identification.

As between-speaker differences are mainly caused by organic differences (i.e. different shapes and sizes of the vocal tract) and learned differences (i.e. different sources from which the speakers learn the language) [1], if two speakers share the same organic structure and the same language environment, the differences between them represent the lower limit of possible between-speaker differences. Identical twins turn out to be such subjects. If the differences between identical twins in certain parameters are large enough to tell them apart, it is reasonable to expect that those parameters should be able to discriminate irrelevant speakers as well.

Previous studies on the static features (e.g. formant centre frequency) of identical twins show that despite the great similarity between identical

twins, differences were found in most twins [3, 7, 8, 9]. Comparing with static features, dynamic features, such as formant dynamics, show more information about individual speakers as they carry not only the information about the target sounds as static features do, but also the information about the movements between various targets. Previous research shows that formant dynamics are more identity-revealing than vowel centre frequency in speaker identification [2, 4]. McDougall [4] analysed the formant dynamics of English /aik/, and found that all the subjects could be very well discriminated by the formant trajectories. However, the subjects she used were irrelevant male speakers. It remains a question whether similar voices like those of identical twins can also be identified by formant dynamics.

Meanwhile, as bilingualism becomes more and more common in the society, we are interested to know whether bilingual speakers have different scopes of variation in different languages. No study so far has investigated the between-speaker differences of bilingual identical twin speakers.

This study examines the formant dynamics of the diphthong /ua/ in Mandarin-Shanghainese bilingual identical twins' speech. The similarities and differences between twins in each language are discussed and assessed by Discriminant Analysis.

2. METHOD

2.1. Subjects

The subjects were eight pairs of identical twins, four pairs of male twins (denoted MT1A, MT1B, MT2A, MT2B, ...) and four pairs of female twins (denoted FT1A, FT1B, FT2A, FT2B, ...), aged 15-26. All of them were born and raised in Shanghai, using both Shanghainese and Mandarin in daily life. Six out of eight pairs of twins lived together since they were born, and shared at least some education background. MT4 and FT4 were raised up separately – MT4A and FT4A were brought up by grandparents on their mothers' side; MT4B and

FT4B were brought up by grandparents on their fathers' side. But these two pairs of twins stayed together during the weekends and communicated with each other frequently.

2.2. Questionnaire

Before the recording session, a questionnaire was given to the subjects, asking about their shared education background, their attitudes towards being twins, and their language use at various settings.

The questionnaire showed that all the female twins and two pairs of male twins (MT1 and MT3) were Mandarin-dominant. They spoke Shanghainese only to their family members, and communicated with their twin siblings mostly in Mandarin. Two pairs of male twins (MT2 and MT4) were Shanghainese-dominant. They used Shanghainese both at work and at home, and communicating with their twin brothers mostly in Shanghainese.

All the subjects reported that their voices had been mistaken for their twin siblings from time to time. FT2 said that their voices were frequently misidentified even by their parents in daily life.

As for their attitudes towards being twins, FT2, FT3 and MT3 found it amusing that other people often got their identity wrong, and they wanted to be the same despite the inconvenience it had brought them. The other five pairs felt indifferent when being mistaken.

2.3. Materials

The subjects were recorded reading the materials six times. The target words consisted of three syllables that are phonemically the same in Shanghainese and in Mandarin. There are in total 36 tokens (3 syllables \times 6 repetitions \times 2 languages). All the target words contained the diphthong /ua/ with a falling tone, a tone common to both languages. The target words are listed below:

- Shanghainese: 乖 /kua/, 夸 /k^hua/, 歪 /hua/
- Mandarin: 掛 /kua/, 跨 /k^hua/, 畫 /hua/

The target words and filler items were embedded in carrier phrases:

- Shanghainese: /ŋu do? ____ gə? ə? zi/.
我讀__個個字。(I read ____ this word)
- Mandarin: /wo tu ____ tɕɿ kɿ tsi/.
我讀__這個字。(I read ____ this word)

2.4. Procedures

The subjects were given some time to practice before the actual recording. The recordings were taken in a quiet room with a Samson Zoom H2 Handy Recorder at a sampling rate of 44100Hz. Each subject was seated in front of the desk with the recorder placed about 20cm from his or her mouth. They were recorded reading the materials in Shanghainese first, and then in Mandarin. The subjects were asked to read the materials in a clear manner with a normal speech rate.

2.5. Measurements

The recordings were downsampled to 22050Hz and analysed using Praat. The beginning and the end of /ua/ were marked manually. The total duration of the vowel was divided into 10 equal parts and F1-F3 frequencies were tracked at each +10% step using a Praat script. The results were checked manually using FFT spectral slices.

2.6. Discriminant Analysis

Discriminant Analysis (DA) was done to assess the similarities of the twins. 15 predictors (F1-F3 frequencies at 10%, 30%, 50%, 70% and 90% points) were used in this study.

3. RESULTS

The results show that although being very similar, the formant dynamics of identical twins still exhibit differences. T-tests reveal that all pairs of subjects have significant differences on some of the 30 points that have been measured. They are not reported in detail here due to page limit. In DA, the overall correct classification rates in each sex and language turned out to be very high (see **Table 1**). The classification rates were calculated separately for female (FT) and male (MT) subjects, and for Shanghainese (SH) and Mandarin (MA) materials. All the tests scored above 80%. This shows that the differences between the formant dynamics of identical twins are large enough to separate them in most situations.

Table 1: Correct identification rate.

	Shanghainese	Mandarin
FT	92.3%	82.5%
MT	88.1%	81.3%

Previous literature [3, 5-7] suggests that higher formants tend to be more identity-revealing. In the current study, however, due to the same organic structure, higher formants of identical twins turned

out to be identical in some cases. FT2's Mandarin, FT4's Mandarin, MT3's Mandarin and Shanghainese, and MT4's Mandarin and Shanghainese show no statistically significant difference on any of the points on F3, while their F1 and F2 do exhibit some significant differences.

One interesting finding of this study is that all the twins show more similarities in their dominant language. MT2A and MT4A are more similar to their twin siblings when they speak Shanghainese, and all the other pairs are more similar when they speak Mandarin. Take MT4 as an example. Their Shanghainese and Mandarin /ua/ show similar scope of variation in F1 and F3 frequencies – no significant between-speaker difference is found in their F3 frequencies in either language, and the F1 centre frequencies show significant between-speaker differences in both languages. However, their F2 frequencies of the two languages exhibit different variability. Table 2 shows whether MT4's F2 of Mandarin and Shanghainese /ua/ are statistically significant different from each other's on 10%-90% points. The asterisk indicates a significant result ($p < 0.05$). It is clear that their Mandarin shows larger between-speaker differences.

Table 2: T-test results of MT4A's and MT4B's F2 of Mandarin and Shanghainese /ua/.

	10	20	30	40	50	60	70	80	90
SH									
MA	*	*	*	*	*	*	*	*	

Discriminant Analysis confirms this observation. Table 3 shows the rates of each subject being misidentified as his or her twin sibling in DA, female on the left, male on the right, Mandarin first, and Shanghainese second. The cells with gray background are the subjects' non-dominant language.

Table 3: Rate of each subject being misidentified as the twin sibling.

	FT (MA)	FT (SH)	MT (MA)	MT (SH)
1A	16.7	11.1	.0	.0
1B	.0	5.9	.0	.0
2A	16.7	11.1	.0	.0
2B	5.6	.0	.0	5.6
3A	5.9	.0	16.7	.0
3B	5.6	.0	11.1	.0
4A	33.3	.0	11.1	38.9
4B	27.8	5.6	.0	22.2

Almost all the non-dominant language data have a lower chance of being misclassified, i.e.

with fewer differences. This is a strong indication that the subjects differ more from their twin in their non-dominant language. A possible account for this phenomenon is that the twins tend to be more similar in the language they use to communicate with each other, as they can assimilate to each other in daily communication. For the non-dominant language, since they seldom communicate with each other using that language, there is little chance for assimilation no matter how uniform the language environment is.

Another interesting finding of this study is that separated twins are not necessarily more different than non-separated twins in terms of formant dynamics. In fact, the two pairs of separated twin subjects are more similar than some of the non-separated twins in the current study. Figure 1 and Figure 2 show the formant trajectories of FT1's (non-separated) and FT4's (separated) Mandarin /ua/ respectively. These two pairs are Mandarin-dominant twins. The figure shows that FT1's (non-separated) average F2s are far apart. In contrast, all the three formants of FT4 (separated) almost overlap with each other.

Figure 1: Average F1-F3 frequencies of FT1's Mandarin /ua/.

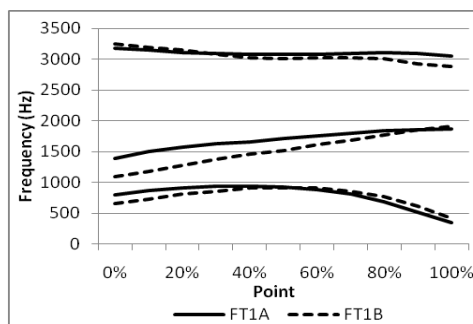
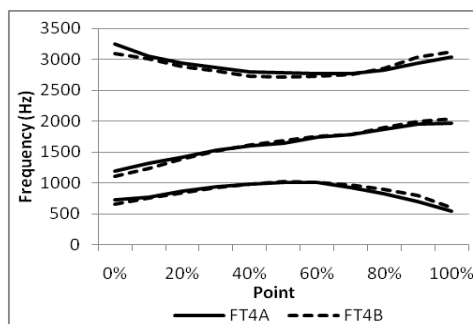


Figure 2: Average F1-F3 frequencies of FT4's Mandarin /ua/.



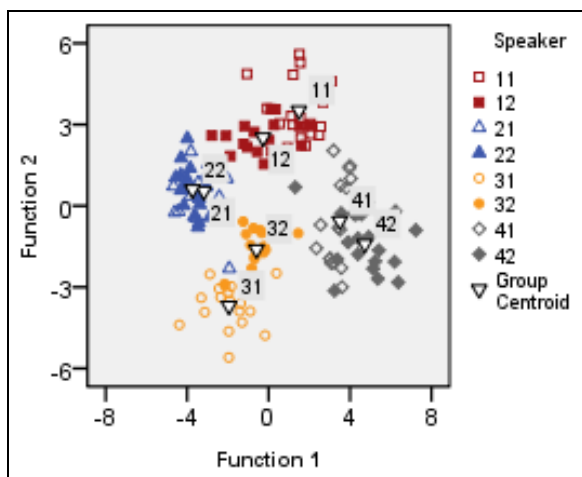
The statistics in Table 3 also suggest that FT4 are more difficult to classify than FT1. 33.3% of FT4A's productions were misidentified as FT4B's, and 27.8% of FT4B's were misidentified as

FT4A's. Comparing with FT4, the misidentifying rates of FT1 are much lower.

Thus, both the visual inspection and statistical analysis confirm that FT4 are more similar than FT1, which means that separated twins are not necessarily more different than non-separated twins. Therefore, given enough interaction between the twins, it is possible that the uniformity of language environment is not the most important in shaping their speech patterns. Individual choice plays a more important role. In other words, identical twins can choose whether to be the same or different irrespective of their language environment.

The effect of individual choice can also be partly revealed by the fact that the three pairs who claimed that they wanted to be the same in the questionnaire (i.e. FT2, FT3 and MT3) tend to show fewer between-speaker differences than other pairs. Visual inspection shows that when Mandarin is analysed, FT2 turned out to be the most similar pair in female twins, and MT3 turned out to be the most similar pairs in male twins. Discriminant Analysis confirms this observation. Figure 3 is the scatter plot of the Canonical DA Function of female twins' Mandarin /ua/. The group centroids of FT2A and FT2B (denoted 21 and 22 in the plot) almost overlap with each other.

Figure 3: Canonical Discriminant Functions of Female Twins' Mandarin /ua/.



It is worth noticing that FT2 were 21-year-old when the recordings were made. They were the oldest pair among the female twins. They had been studying in different schools and universities for many years and had different social networks. In other words, their language environments were largely different from each other's. Despite these

differences, they showed extremely similar formant dynamic patterns.

Therefore, the high similarity between the twins who wish to sound the same may indicate that individual choice plays a more important role in one's speech patterns than organic structure.

4. CONCLUSION

The data in this study shows that although being very similar, the twins' formant dynamics still exhibit some differences which can be used to distinguish them. Therefore, uniform organic structure and language environment does not necessarily result in the same voice. The facts that separated pairs are not necessarily more different than non-separated pairs and that the twins are more similar in their dominant languages in spite of a more uniform language environment of their non-dominant languages suggest that individual choices play a more important role than the language environment in shaping the voices of identical twins. Identical twins can have very similar formant dynamics if they wish to sound the same. However, a more accurate way of assessing and quantifying the subjects' willingness to be twins is needed in further studies.

5. REFERENCES

- [1] Garvin, P.L., Ladefoged, P. 1963. Speaker identification and message identification in speech recognition. *Phonetica* 9, 193-199.
- [2] Greisbach, R., Esser, O., Weinstock, C. 1995. Speaker identification by formant contour. In Braun, A., Köster, J. (eds.), *Studies in Forensic Phonetics: Beiträge zur Phonetik und Linguistik*. Trier: Wissenschaftlicher Verlag Trier.
- [3] Loakes, D. 2006. *A Forensic Phonetic Investigation into the Speech Patterns of Identical and Non-Identical Twins*. PhD dissertation, University of Melbourne.
- [4] McDougall, K. 2005. *The Role of Formant Dynamics in Determining Speaker Identity*. PhD dissertation, University of Cambridge.
- [5] Nolan, F. 1983. *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.
- [6] Nolan, F. 1997. Speaker recognition and forensic phonetics. In Hardcastle, W., Laver, J. (eds.), *The Handbook of Phonetic Sciences*. Oxford: Blackwell, 744-767.
- [7] Nolan, F., Oh, T. 1996. Identical twins, different voices. *Forensic Linguistics* 3, 39-49.
- [8] Whiteside, S.P., Rixon, E. 2001. Speech patterns of monozygotic twins: An acoustic case study of monosyllabic words. *The Phonetician* 82, 9-22.
- [9] Whiteside, S.P., Rixon, E. 2003. Speech characteristics of monozygotic twins and a same-sex sibling: An acoustic case study of coarticulation patterns in read speech. *Phonetica* 60, 273-297.