Formant Dynamics of Bilingual Identical Twins in Non-Contemporaneous Speech

Donghui Zuo and Peggy Mok

Department of Linguistics and Modern Languages, The Chinese University of Hong Kong
donghui@cuhk.edu.hk, peggymok@cuhk.edu.hk

Abstract

This study investigates the formant dynamics of Mandarin and Shanghainese /ua/ in bilingual identical twins. Data collected from two sessions were analysed using Discriminant Analysis. The results show that despite the same organic structure and similar language environment, differences can be found in the formant dynamics of identical twins. However, cross-session analysis suggests that for identical twins, long-term within-speaker differences can be larger than between-speaker differences. A comparison between the subjects' dominant and non-dominant language shows that the twins are more similar in their dominant language. We also found that separated twins are not necessarily more different than non-separated twins.

Index Terms: speaker identification, formant dynamics, identical twins, bilingual, non-contemporaneous speech.

1. Introduction

Voice is identity-revealing. Most people are able to recognise familiar people from their voices. However, voice is not unique. Misrecognitions occur from time to time. Whether there are speakers being identical in certain acoustic dimensions is still unclear. To address this issue, it is important to study the lower limit of between-speaker differences and see whether such differences can be smaller than within-speaker differences. Identical twins are ideal subjects for such studies. As between-speaker differences are mainly caused by organic differences (i.e., different shapes and sizes of the vocal tracts) and learned differences (i.e., different sources from which the speakers learn the language) [1], the difference between identical twins is always assumed to be the lower limit of possible between-speaker difference because they have very similar organic structures and language environment.

Previous studies on static acoustic features (e.g. formant centre frequency) of identical twins show that despite the great similarity, differences could be found in most identical twins [3, 4, 7]. Comparing with static features, dynamic features, such as formant trajectories, show more information about individual speakers, as they carry not only the information about the target sounds, 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, 5]. McDougall [5] analysed the formant dynamics of /aɪk/ in English, and found that all the subjects could be well discriminated by their formant trajectories. However, the subjects she used were unrelated speakers. It remains a question whether similar voices like those of identical twins can also be identified by formant dynamics.

In addition, acoustic features change over time. It was found that the discriminatory power of formants would decrease when data from different recording sessions were involved [e.g. 8]. Therefore, to further explore the discriminatory power of formant dynamics, it is necessary to compare between- and within-speaker differences across different sessions.

Meanwhile, as bilingualism is very common in many parts of the world, we are interested to know whether bilingual speakers show different degrees of variation in different languages.

The current study examined the formant dynamics of the diphthong /ua/ in the speech of Mandarin-Shanghainese bilingual identical twins. Non-contemporaneous data collected from two recording sessions were analysed. The similarities and differences between twins are discussed and assessed by Discriminant Analysis.

2. Method

2.1. Subjects

The subjects consisted of eight pairs of identical twins aged 15-26. There were four pairs of male twins (denoted MT1A, MT1B, MT2A, MT2B, and so on) and four pairs of female twins (denoted FT1A, FT1B, FT2A, FT2B, and so on). All of them were born and raised in Shanghai, using both Shanghainese and Mandarin in daily life. Six out of eight pairs (FT1, FT2, FT3, MT1, MT2 and MT3) were non-separated twins, i.e., they had been living together since birth and shared at least some education. The remaining two pairs (MT4 and FT4) were separated twins, i.e., they were raised 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 separated twins stayed together during weekends and communicated with each other by telephone frequently on weekdays.

2.2. Questionnaire

Before the first recording session, a questionnaire was given to the subjects, asking about their shared educational background, attitudes towards being twins, and language use in 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 communicated with their twin siblings mostly in Shanghainese.

All the subjects reported that their voices had been mistaken for their twin siblings from time to time in at least one of the two languages they use. FT2 said that their voices had been 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 were indifferent when being mistaken.

2.3. Materials

The diphthong /ua/ was chosen as the target sound for two reasons. First, this sound can be found in both languages, which allows us to compare across languages. Second, the degree of movement in /ua/, especially in F2, is quite small. As a result, this sound tends to yield poorer identification results than sounds that involve greater movements. Hence, it is of interest to us whether similar-sounding speakers can be discriminated using this sound.

Two word lists (one in Shanghainese and one in Mandarin) both containing target words and filler items were used in this study. The target words were three syllables /kua/, /k^hua/, /hua/ that are phonetically similar in Shanghainese and in Mandarin (Shanghainese stimuli: 乖 'obedient', 誇 'praise', 歪 'skewed'. Mandarin stimuli: 掛 'hang', 跨 'cross', 畫 'draw'). All the target words had a falling tone, a tone common to both languages. They were all common words in the two languages. The target words and filler items were randomised in order to prevent the subjects from knowing the aim of the study. All the words were embedded in carrier phrases which had the same meaning in the two languages:

- Shanghainese: /ŋu doʔ___ gəʔ ɦəʔ ziɿ/. (I read ____ this word).
- Mandarin: /wo tu ___ tʂɤ kɤ tsiɿ/. (I read ____ this word).

2.4. Procedures

The subjects were recorded reading the same word lists in two sessions separated by at least one and a half months (the interval ranged from 1.5 months to 7 months). All the recordings were taken in a quiet room with a solid state recorder at a sampling rate of 44100Hz.

At the beginning of each session, the subjects were given some time to practise before the actual recording. When the recording began, they were seated in front of a desk with the microphone placed about 20cm from his or her mouth, and were asked to read the Shanghainese list six times and then the Mandarin list six times in a clear manner with a normal speech rate. Short breaks were given between lists. 576 tokens in total (3 syllables × 6 repetitions × 2 languages × 16 speakers) were collected from the recordings.

2.5. Measurements

The recordings were downsampled to 22050Hz and analysed using Praat. The beginning of F1 and the end of F2 of /ua/, which were considered the beginning and the end of the vowels, were marked manually. The total duration of the vowel was divided into 10 equal intervals 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 (DA)

Discriminant Analysis was performed to assess the similarities of the twins. 15 predicators (F1-F3 frequencies at 10%, 30%, 50%, 70% and 90% points) were used in this study. Outliers were excluded from the datasets. The classification rates were calculated for female (FT) and male (MT) subjects, and for Shanghainese (SH) and Mandarin (MA) materials separately. Each token was classified using Leave-on-out classification (cross-validation).

3. Results and discussions

The results suggested that despite the identical organic features and shared language environments, the twins did exhibit differences in their formant dynamics when the data were collapsed across sessions. MANOVA showed that all the 8 pairs of subjects had significant differences in their formant dynamics when the 27 points on the first three formants were considered together (due to the page limit, the results are not reported in detail here). Besides MANOVA, the high overall correct classification rates from Discriminant Analysis in each sex and language also show that the twins can be discriminated by their formant dynamics (see Table 1). Even for the female Mandarin data, which yielded the lowest classification rates, 77.8% of all the tokens were correctly classified. This indicates that the identical twins do have larger between-speaker difference than within-speaker difference in their formant trajectories.

Table 1: Correct classification rates in Discriminant Analysis (FT stands for female twins, and MT stands for male twins).

	Shanghainese	Mandarin
FT	90.0%	77.8%
MT	81.5%	78.6%

Table 2 illustrates the rates of each subject being misidentified as his or her twin sibling in DA. The higher the number, the more that specific subject resembled his or her twin sibling. The cells with gray background are the subjects' dominant languages.

Table 2: Rate of each subject being misidentified as their twin sibling (MA stands for Mandarin, and SH stands for Shanghainese).

	FT (MA)	FT (SH)	MT (MA)	MT (SH)
1A	13.9	5.7	8.3	3.0
1B	22.2	12.9	11.1	5.6
2A	14.3	16.7	2.8	13.9
2B	8.3	11.4	.0	2.8
3A	17.1	.0	29.4	14.7
3B	20.0	5.6	28.6	14.7
4A	28.6	.0	8.3	16.7
4B	25.0	2.8	8.3	16.7

One interesting finding of this study is that all the twins exhibited more similarities in their dominant language in terms of the rates of being misidentified as their twin siblings. All the twins except FT2 showed higher chances of being misclassified in their dominant language (i.e., MT2 and MT4 had higher misidentification rates in Shanghainese, and others had higher misidentification rates in Mandarin). This is a

strong indication that the subjects tend to differ more from their twin siblings in their non-dominant language. A possible account for this phenomenon is that, as the twins communicate with each other in their dominant language, frequent daily conversations provide them with more chances to assimilate to each other. As for the non-dominant language, since they seldom talk to each other using that language, there is little chance for assimilation no matter how uniform the language environment is. In fact, this account is also compatible with FT2's pattern. While other Mandarin-dominant pairs claimed that they seldom communicate with their twin siblings in Shanghainese, FT2 reported that although Mandarin was their dominant language, they still used Shanghainese in daily conversation quite often. In other words, FT2 was more balanced than the other pairs in terms of language dominance. As a result, it is not surprising that their misidentification rates were comparable across languages.

Another 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 appeared to be more similar than some of the non-separated twins in the current study. Figure 1 and 2 show the formant trajectories of MT2's (non-separated) and MT4's (separated) Shanghainese /ua/ respectively. These two pairs were both Shanghainese-dominant twins and were of the same age. MT4 had been living separately for over 22 years while MT2 had been living together all the time. However, contrary to our expectation, the figures clearly show that the first three formants of MT2 (non-separated), especially F3, show distinct patterns. In contrast, all the three formants of MT4 (separated) almost overlap completely.

Figure 1 Average F1-F3 of MT2's Shanghainese /ua/.

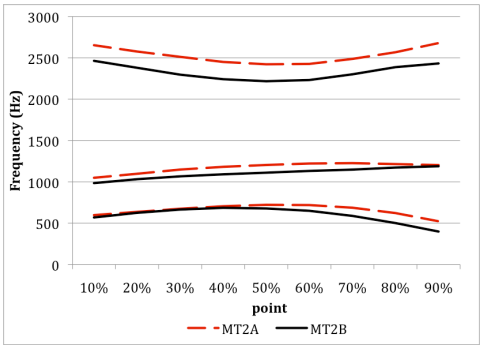
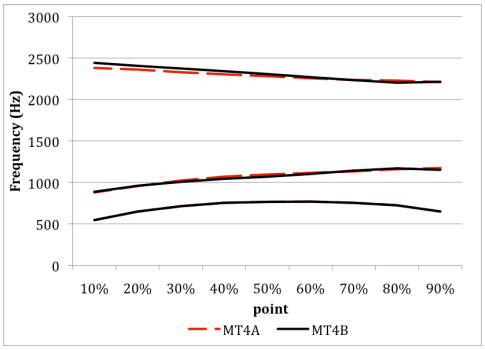


Figure 2 Average F1-F3 of MT4's Shanghainese /ua/.



The misclassification rates in Table 2 also suggest that MT4 are more difficult to classify than MT2. 16.7% of MT4A's productions were misidentified as MT4B's, and 16.7% of MT4B's were misidentified as MT4A's. Comparing with MT4, the misidentifying rates of MT2 were lower, especially for MT2B whose Shanghainese /ua/ were misidentified as MT2A's in only 2.8% of occasions.

Thus, both visual inspection and statistical analysis suggest that MT4 were more similar than MT2. This means that separated twins can be more similar in formant dynamics 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 factor in shaping their speech patterns. Individual choice plays a more important role. Identical twins can choose to be the same or different irrespective of their language environment. The above two findings confirm the results we got from the study based on contemporaneous recordings of the same pairs of twins [9].

Besides using absolute formant frequency values, we also characterising the first three formant contours with three cubic polynomial equations (i.e., each formant was fitted with an equation $y = a_0 + a_1t + a_2t^2 + a_3t^3$) and used the coefficients (i.e., a_0, a_1, a_2, and a_3 of the three equations, 12 predictors in all) to conduct classifications (following the method in [6]). Both the classification rates and the patterns were similar to the analysis using the absolute values. Table 3 shows the correct classification rates in Discriminant Analysis using cubic polynomial coefficients. The figures are comparable to those in Table 1. In fact, except for female Mandarin, all the three conditions yielded even slightly lower classification rates when coefficients were used. Such results may be accounted by the fact that the cubic polynomial equations are approximations of the formant dynamics instead of the actual dynamics. Although the R values (i.e., how good the dynamics are fitted with the equations) were quite high for each trajectories (all over 0.8), some subtle information was filtered out. Such information may be important in identifying speakers like identical twins, who can be very similar in their dynamic acoustic features. Therefore, we decided to use the absolute values instead of the polynomial equations in further analysis.

Table 3: Correct classification rates in Discriminant Analysis using the cubic polynomial coeffients of the first three formants (FT stands for female twins, and MT stands for male twins).

	Shanghainese	Mandarin
FT	88.4%	79.2%
MT	80.4%	78.1%

In the discussion above, the data were analysed collapsed across sessions (i.e., the Discriminant functions were calculated based on the tokens collected from both sessions). However, to see whether between-speaker differences are always larger than within-speaker differences, we also need to compare the long-term between-speaker differences with within-speaker differences. To maximise within-speaker differences, we separated the data into two parts – the recordings from the first session were treated as a training set, and the recordings from the second session were treated as a test set. Discriminant functions were calculated based on the training set, and classification was conducted for the test set. The cross-session classification rates are shown in Table 4.

Table 4: Cross-session correct classification rates.

	Shanghainese	Mandarin
FT	74.8%	61.5%
MT	45.3%	59.2%

Comparing with Table 1, the results in Table 4 show that when long-term between- and within-speaker differences are taken into account, the cross-session classification rates drop considerably. The highest classification rate was found in female subjects' Shanghainese /ua/, while the classification rate of male subjects' Shanghainese /ua/ was below 50%. Careful examination of the classifications suggested that most misidentifications involved identical twins. Table 5 illustrates the cross-session misidentification rate of each subject being misidentifies as their twin sibling.

Table 5: Cross-session misidentification rate of each subject being misidentified as their twin sibling**.**

	FT (MA)	FT (SH)	MT (MA)	MT (SH)
1A	5.6	41.2	.0	.0
1B	38.9	14.3	22.2	77.8
2A	44.0	5.6	16.7	5.6
2B	5.6	22.2	.0	.0
3A	11.8	5.6	43.8	56.3
3B	73.7	.0	83.3	55.6
4A	11.8	11.1	0	.0
4B	55.6	5.6	16.7	83.3

It is surprising to see that over half of FT3B's, FT4B's and MT3B's Mandarin /ua/ and MT1B's, MT3A's, MT3B's and MT4B's Shanghainese /ua/ were misclassified as their twin siblings'. The misclassification rated of MT3B's Mandarin and MT4B's Shanghainese were even higher than 80%. In other words, their formant dynamics in session 2 resembled those of their twin sibling's more than those of themselves in session 1. The high misclassification rates suggest that for identical twins, long-term within-speaker differences can be larger than short-term between-speaker differences.

One possible reason for such large discrepancies between within-session and cross-session results might be caused by the nature of the word list reading task and the sensitivity to variance of Discriminant Analysis. In reading task, the subjects tend to maintain their style and speech rate within one session. That is to say, the within-speaker variations are minimal within each session. However, in this study, the subjects came back after more than one month. Although they were told to read with a similar manner to that they used in Session 1, it was difficult for all the speakers to recall the exact style. Any small differences can cause the cross-session within-speaker difference larger than within-session difference. Since Discriminant Analysis is relatively sensitive to variances, higher within-speaker differences would cause more overlapping between members of identical twins. The difference in variance is likely to be the major cause of such high misclassification rate, because when the data from the two recording sessions were collapsed (i.e. the long term within-speaker differences were taken into consideration when the Discriminant functions were calculated), the classification rate turned out to be very high. To further investigate the similarity and differences between the speech patterns of identical twins, it is necessary to look into spontaneous speech, as the within-speaker differences in it would be larger than those in list reading tasks. Alternatively, we can try other statistical tests which are less sensitive to different variances to see whether non-contemporaneous speech of identical twins can be correctly identified.

4. Conclusions

The data in this study show that although being very similar, the twins' formant dynamics still exhibit some differences which were large enough to discriminate them when the data were collapsed across sessions Therefore, uniform organic structure and language environment does not necessarily result in the same voice. The facts that separated twins were not necessarily more different than non-separated twins and the fact that the twins were 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.

On the other hand, although Discriminant Analysis was found to be successful in classifying data when the recordings collected from two sessions were analysed together, cross-session analysis showed that many identical twins do have larger long-term within-speaker differences than between-speaker differences.

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] Nolan, F., Oh, T. 1996. Identical twins, different voices. Forensic Linguistics 3, 39-49.

[5] McDougall, K. (2005). The Role of Formant Dynamics in Determining Speaker Identity. PhD dissertation. University of Cambridge.

[6] McDougall, K. (2006). Dynamic Features of Speech and the Characterisation of Speakers: Towards a New Approach Using Formant Frequencies. International Journal of Speech, Language and the Law 13 (1), 89-126.

[7] Whiteside, S.P., Rixon, E. 2001. Speech Patterns of Monozygotic Twins: An Acoustic Case Study of Monosyllabic Words. The Phonetician 82, 9-22.

[8] Rose, P., Clermont, P. 2001. A Comparison of Two Acoustic Methods for Forensic Speaker Discrimination. Acoustics Australia 29 (1), 31-35.

[9] Zuo, D., Mok, P. 2011. Formant dynamics of /ua/ in the speech of Mandarin-Shanghainese bilingual identical twins. In Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS), 2332-2335.