Journal of Phonetics 52 (2015) 1–12

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

Formant dynamics of bilingual identical twins

Donghui Zuo, Peggy Pik Ki Mok *

Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Shatin, Hong Kong

ARTICLE INFO

Article history:
Received 6 March 2014
Received in revised form
20 March 2015
Accepted 26 March 2015
Available online

Keywords:
Formant dynamics
Diphthongs
Identical twins
Bilingual
Chinese

ABSTRACT

This study investigates speech similarities in bilingual identical twins using formant dynamics. Previous studies show both similarities and differences in static features between twin speakers, but it is unclear if differences also exist in dynamic features, given that identical twins have the same vocal tract anatomy which allows them to reach the same phonetic targets with the same articulatory strategies. Moreover, previous studies show that bilingual speakers exhibit different speech patterns in their two languages. The degree of variability between the two languages of identical twins is also unknown. Eight pairs of Shanghainese-Mandarin bilingual identical twins were recorded producing the common diphthong /ua/ in both languages in two recording sessions. Frequencies of the first four formants were measured at each +10% step. The results show clear differences in formant dynamics between identical twins, and the differences were large enough to discriminate them using Discriminant Analysis. The twin speakers were more similar in their dominant language than in their non-dominant language. Twins raised separately were as similar as twins growing up together. The results suggest that individual choices play an important role in shaping the speech of identical twins.

1. Introduction

This study investigates the speech similarities in bilingual identical twins using formant dynamics in diphthongs. Identical twins can be considered the most similar speakers in a population, and anecdotal observations suggest that many identical twins do sound quite similar. Previous studies have shown that formant dynamics have good discriminating power in speaker identification (e.g. Jessen, 1997; Kinoshita, 2001; Loakes, 2006). The current paper seeks to determine if formant dynamics can also successfully discriminate identical twin speakers. In addition, the speech similarities of identical twins can be further evaluated using twins who are bilingual as many studies have demonstrated that bilingual speakers can have different speech patterns in their two languages (e.g. Fowler, Sramko, Ostry, Rowland, & Hallé, 2008; Guion, 2003; Macleod & Stoel-Gammon, 2005). The results of the study can shed some light on issues concerning speech production in identical twins.

1.1. Speech patterns of identical twins

A widely accepted model of between-speaker differences (e.g. Garvin & Ladefoged, 1963; Glenn & Kleiner, 1968; Wolf, 1972) claims that any type of between-speaker difference is either organic (i.e., caused by different anatomical structures of the vocal tract) or learned (caused by different articulatory manners acquired by speakers). In this sense, identical twins who stay together would have identical voices, as they are identical in anatomy and also share the same language environment.

In terms of anatomy, identical twins are assumed to have the same shape and size of vocal tracts, since genetic information plays a major role in determining a person's overall size, shape and rate of growth and maturation (Beck, 1997). Besides the overall shape and size of the vocal tracts, their speech is also influenced by the same dentition problems. Townsend, Richards, Hughes, Pinkerton, and Schwerdt (2003) found that dentition problems like tooth caries, periodontitis, and malocclusions all have genetic bases. Minor differences may exist, but they are considered as small as the “left-right asymmetries” existing in one individual (Lundström, 1948). Moreover, Thompson et al. (2001) found that the frontal and language-related cortices as well as the quantity of frontal grey matter are highly similar in identical twins. Therefore, the anatomical variations between identical twins are negligible.

* Corresponding author. Tel.: +852 39431768.
E-mail address: peggymok@cuhk.edu.hk (P.P.K. Mok).

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.03.003

In terms of language environment, previous studies have revealed its importance in twin speech. Mayr, Price, and Mennen (2012) found that migrating to an L2-speaking country changed a speaker's VOT and vowel quality comparing to her identical twin sister who stayed in her L1-speaking country. In contrast, identical twins growing up together are exposed to similar speaker groups. They have the same close relatives who play major roles in their language acquisition. Therefore, the language environments of these twins can be regarded as controlled as well.

With identical anatomy and highly similar language environments, identical twins exhibit high similarity in many acoustic dimensions. Van Gysel, Vercammen, and Debruyne (2001), Przybyla, Horii, and Crawford (1992) and Debruyne, Decoster, Van Gijsel, and Vercammen (2002) analysed the fundamental frequency of identical twins, and found that identical twins have more similar fundamental frequency values than unrelated speakers. Van Lierde, Vinck, De Ley, Clement, and Van Cauwenberge (2005) measured the maximum phonation time, highest frequency, lowest intensity and $F0$ of identical twins, and found no significant differences between identical twins.

However, differences in anatomy and language environment are not the only sources of between-speaker variations. Nolan (1997) pointed out that there may be some aspects that speakers are free to choose in speech production. They are not part of the shared language systems, nor strictly constrained by the vocal mechanisms. These differences are termed as "choices" (Brown, 1987:17; Nolan & Oh, 1996:40; Rose, 2002:188). Choices can be conscious or subconscious. For conscious choices, speakers can actively decide whose speech patterns to adopt. For subconscious choices, it is possible that they are mediated by social and linguistic factors (Babel, 2009). Some free random variations may also be involved.

Indeed, acoustic differences also exist in the speech of identical twins. Loakes (2006, 2008) studied the formant centre frequencies in the speech of one pair of non-identical twins and three pairs of identical twins. All the subjects had lived their life together by the time of recording. She found that only one identical twin pair showed no significant difference in their vowel centre frequencies, while the other three pairs all had significant differences. Besides, Whiteside and Rixon (2001) studied a pair of identical twins who had not been separated. Their average formant frequencies showed that one subject had a consistently higher $F1$ mid value and a lower $F2$ midvalue than his twin sibling. These results indicate that besides anatomy and language environment, individual choice (e.g. what they want to sound like and the impression they would like to make on others) also plays an important role in the speech production of identical twins.

In fact, both similarities and differences can co-exist in the speech of identical twins. Nolan and Oh (1996) studied the syllable-initial /l/ and /r/ of three pairs of identical twins. One pair of twins realised /r/ differently and showed differences in the $F3$ of the vowels following /r/. One pair differed in the realisation of /l/. The third pair differed in the $F2$ of their /uː/. One surprising finding of that study is that, although consistent differences were observed in all the three pairs of twins, there was no strong difference in coarticulation between the twins. Thus, Nolan and Oh proposed that since identical twins have the same anatomy, they can produce the same target sound in the same way so that their coarticulation is extremely similar. Similarly, Whiteside and Rixon (2003) assessed the degree of coarticulation in the speech of identical twins, and found that the twins used very similar articulatory gestures at the onset of vowels. Nonetheless, single-point measurement cannot adequately capture dynamic coarticulatory patterns. To investigate the rate of change of vocal tract configurations, formant contours are better parameters than single-point measurement.

The aforementioned studies suggest that different acoustic parameters in the speech of identical twins are not affected equally. Weirich (2011, 2012) investigated four pairs of identical twins and three pairs of non-identical twins with both acoustic and articulatory data. Weirich found that identical twins exhibited higher similarities than non-identical twins in some parameters (e.g. the $F1$/$F2$ dimensions of vowel spaces), while in other parameters (e.g. the formants centre frequency of stressed /iː/, /a/, and /uː/, the spectral mean value and the highest amplitude of /s/), identical twins are not necessarily more similar than non-identical twins. What is of particular interest to the present study is that, Weirich found that identical twins show greater similarities in all the dynamic features she studied (i.e., the sibilant–vowel transitions and the /aka/ transitions), but not necessarily in the static features (e.g. the formant centre frequency of stressed /iː/, /a/, and /uː/). So Weirich concluded that static targets are learned features, while transitions are determined by anatomy (cf. Kühnert & Nolan, 1999; Nolan, Oh, McDougall, De Jong, & Hudson, 2006; Rose, 2002). However, Weirich studied the /aka/ transitions using articulatory measures, while for the sibilant–vowel transitions, she quantified the transition using only the differences between the formant frequency values at the end of the sibilant and the formant frequency values of the target vowel at 40 ms after the sibilants. Such acoustic measures may oversimplify the dynamic features in speech, as they cannot capture the rate of change in the formants across an utterance. Therefore, in order to better examine whether transitions between different sound targets are determined by anatomy and whether identical twins differ in this dimension, finer measurements are needed.

1.2. Formant dynamics

Many studies have shown that formant frequencies can discriminate speakers well (e.g. Jessen, 1997 on German /a/; Kinoshita, 2001 on Japanese vowels; Loakes, 2006 on Australian English vowels). In these studies, the centre or onset formant frequencies of vowels were measured, i.e., they all used single-point measurement and investigated only the static features of the vowel frequencies. In the present study, formant dynamics of diphthongs are chosen as they reflect both the static features of the formants at multiple points and the dynamic features of how the formants move from one target sound to another.

Formant dynamics refer to changes in formant frequencies over time. These time-varying spectral features of vowels are also called vowel inherent spectral changes (Nearey & Assmann, 1986). Different factors contribute to these spectral changes, for example, vowel-specific formant trajectories, consonantal contexts and prosodic effects (e.g. emphatic stress). Dialects of the same language can also differ in formant dynamics (Fox & Jacewicz, 2009). Although the primary research focus of these dynamic formant changes is their relevance in vowel perception (e.g. see Morrison & Assmann, 2013), these vowel inherent features, especially vowel-specific trajectories like those in diphthongs, allow much freedom for speaker-specific behaviours in production, e.g. variability in the amount of spectral change and the rate of change. Even if speakers produce the same acoustic targets, they can still differ in the transitions between targets. Therefore, formant dynamics are good candidates to compare between-speaker differences in speech production of identical twins.

Formant dynamics have stronger discriminatory power than vowel centre frequencies in both machine speaker recognition and machine speaker identification using statistical procedures. Greisbach, Esser, and Weinstock (1995) compared single-point measurement and multiple-point measurements in machine speaker recognition. They examined the formant trajectories of six German vowels, measuring the $F1$ and $F2$ of each token at every +25% point. It turned out that for all six vowels, the identification rate was much higher when multiple points were used compared to using the centre frequency values only. Ingram, Prandolini, and Ong (1996) also measured different points along the formant trajectories and integrated the information in machine recognition, and the results showed a high recognition rate. Therefore, formant dynamics exhibit more between-speaker differences than vowel centre formant frequencies do.

Several studies have demonstrated the success of using formant dynamics in machine speaker identification. Rose, Kinoshita, and Alderman (2006) compared 25 male-speakers using the diphthong /aɪ/ in Australian English. Formant frequencies were measured at two sampling points – the steady states of /a/ and /ɪ/ identified from the spectrogram. The statistical model yielded a very high discrimination rate. Zhang, Su, Cao, and Zhao (2010), Zhang, Morrison, and Thiruvaran (2011) did a similar study with a set of Mandarin Chinese syllables containing either the diphthong /aɪ/ or the triphthong /ɪao/ produced by 20 female speakers. The formant frequencies at the starting points, the midpoints and the end points were used to perform a speaker identification task. When the values of the three points were combined, the identification rate was over 95%.

Both Rose et al. (2006) and Zhang et al. (2011) used only a few measuring points to represent the formant dynamics of diphthongs. More information from the whole formant trajectories can yield even better identification results. Morrison (2009) investigated 27 speakers' production of five diphthongs /ɔɪ/, /aɪ/, /oʊ/, /aʊ/ and, /eɪ/ in Australian English. He tracked the frequency of $F1$–$F4$ every 2 ms. The correct identifications rates were above 90% for all the five diphthongs. McDougall (2004, 2006) investigated the formant dynamics of /aɪk/ in Australian English in the speech of five male speakers, both at normal and at fast speech rate, with nuclear stress and without nuclear stress. She measured the frequencies of the first three formants at every +10% point. Discriminant Analysis was used to conduct speaker identification. The results showed that with the data at only three measuring points (i.e. 10%, 50%, 90% points), a classification rate of around 90% was achieved in all conditions.

All of the above studies demonstrate the discriminatory power of formant dynamics which carry much speaker-specific information. However, very few studies have taken similar-sounding speakers into consideration. Only one study by Rose (1999) involved similar-sounding speakers. The subjects were six pairs of similar-sounding speakers, most of which were family members. Rose recorded them saying *hello* with different intonation patterns. The measurements were made at the midpoint of the first vowel, the midpoint of /l/ and every +25% point of the second vowel. He calculated the distance between the formants, and found that similar-sounding pairs showed fewer between-speaker differences than randomly sampled speaker pairs. Rose's data indicated that similar-sounding speakers are difficult to discriminate, even using formant dynamics. Given a group of even more similar-sounding subjects – identical twins, it is unclear whether formant dynamics can tell them apart. This is the question that the current study addresses.

1.3. The present study

The present study differs from the previous ones in several aspects. The first and most important difference is the combination of a special group of speakers (i.e., identical twins) and formant dynamics. One possibility is that, given that identical twins have the same vocal tract anatomy, they would reach the same phonetic targets with the same articulatory strategies. If so, the subjects would show very few between-speaker differences in the movement from one target sound to another, and even formant dynamics may not be able to discriminate them correctly. Conversely, it is also possible that since categorical differences have been found at static points in identical twins' speech, more between-speaker differences can be observed in formant dynamics, resulting in good discrimination.

Secondly, the current study makes use of Mandarin-Shanghainese bilingual speakers and compared the between-speaker differences in their dominant and non-dominant languages, while previous twin studies all used monolingual speakers. According to Bradlow (1995), different languages have different bases-of-articulation (i.e., language-specific articulatory or phonetic settings). Many studies have investigated the speech patterns of bilingual speakers and found that bilinguals have different patterns when speaking the two languages. Most of these studies focused on voice onset time (VOT). For example, Fowler et al. (2008) showed that bilinguals have different voiceless stops in English and French, although their patterns of each language differ from the respective monolingual speakers. Macleod and Stoel-Gammon (2005) and Sundara, Polka, and Baum (2006) had very similar findings in their study of English-French bilinguals. All these results confirmed the different bases-of-articulation in bilingual speakers.

In addition to VOT, effects of different languages can be found in other acoustic parameters as well, e.g. vowel duration (Whitworth, 2000) and vowel quality (Guion, 2003). Guion (2003) studied the vowel systems of Quichua-Spanish bilinguals, and found that those bilinguals who acquire two languages before the age of seven have different vowel systems for the two languages they speak.

The vowel systems of the two target languages in the present study differ in their inventories. According to Lee and Zee (2003), there are six monophthongs (/i y aə u ɤ/), eleven diphthongs (/ai au ou uo ei ye ie ia ua uə iu/) and four triphthongs (/iau uai iou uei/) in Mandarin. The vowel inventory of Shanghainese is not as clear because different studies gave slightly different analysis, and they follow the traditional Chinese phonological analysis of listing rhyme units instead of analysing vowels as phonemes. There are at least ten monophthongs (/i y e ø ɛ a ɤ ɔ o u/) and at least eight diphthongs (/ia io ie iu ua uo uɛ uə/) in Shanghainese (Xu & Tang, 1988; Zhu, 1999; Chen, 2008). To the best of our knowledge, there is no study comparing the acoustic realisation of different vowels and diphthongs in Mandarin and Shanghainese. Nevertheless, given the differences in the vowel inventories of the two languages, bilingual speakers might well have differences in the acoustic realisations of the 'same' diphthong /ua/. It is possible that the same diphthong may have different discriminatory power in the two languages, and we are interested to see if such differences are affected by language dominance.

The third difference between the current study and the previous twin studies is that most of the previous studies use only twins that have not been separated, while in the current study, we compared separated and non-separated identical twins: two pairs of identical twins were raised separately from childhood, so their language environments are more heterogeneous than those twins staying together since birth. Ryalls, Shaw, and Simon (2004) investigated the VOT of two pairs of identical twins, one younger pair and one older pair. They found that the VOT averages were similar for the younger twins, but dramatically different for the older twins probably due to several decades of separation. A comparison between the separated and non-separated twins in this study can show how important the language environment is in shaping one's speech patterns.

There are three major research questions in this study. The first one is whether identical twins have similar formant dynamics; and if differences exist, whether they are large enough to discriminate the twins. The second question is whether bilingual twin speakers have different levels of variability in the two languages they speak. The third question is how important the language environment is in shaping one's speech patterns. Since the current study compares speaker differences between separated twins and those between non-separated twins, larger between-speaker differences in separated pairs than in non-separated pairs can be expected if language environment plays a more important role than choices in speaker-specificity.

2. Method

2.1. Subjects

Eight pairs of identical twins (four male and four female pairs) participated in the current study. They are denoted FT1, FT2, FT3 FT4, MT1, MT2, MT3, and MT4 respectively (FT for female Twins, MT for Male Twins). For each pair of twins, one member is denoted A, and the other is denoted B (for example, MT1 pair includes two speakers – MT1A and MT1B).

All the speakers, as well as their parents, were born and raised in Shanghai, China, and had not lived outside Shanghai for more than half a year in their whole life. All the subjects were first screened for auditory similarity based on the first author's impression and were asked whether their voices had often been confused by others. Their Shanghainese and Mandarin proficiency was evaluated using a translation task and a free talking task. All of them were judged as native speakers of both languages by the first author who is a bilingual native speaker of both languages. The background information of the subjects is listed in Table 1.

All the speakers were aged between 15 and 26. Six out of eight pairs had been living together since birth. 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 at weekends and communicated with each other frequently on the phone during weekdays. Therefore, despite the fact that the two pairs of twins were separated from birth, they still had exposure to their twin siblings' speech. None of the subjects reported having any kind of surgery on their vocal tract or dentition. The appearance of each pair of twins was judged to be very similar by the first author.

Besides the basic information, some further background information related to the current study was obtained from the twins using a questionnaire which contained questions about their shared educational background, their language use in various settings (i.e., what language(s) they use in speaking to different groups of people), and their attitudes towards being mistaken as their twin siblings (whether they found it amusing, indifferent or annoying). The results of the questionnaire are summarised in Table 1.

The questionnaire showed that all the non-separated twins as well as a pair of separated twins MT4 (despite the separated childhood) lived together at the time when the recordings were made. Regarding the educational background, except for the

Table 1
Background information of the eight twin pairs.

Speaker	Age	Separated environment	Current Status	Dominant language	Shared education	Frequency of being mistaken as twin sibling	Attitudes towards being mistaken
MT1	24	No	Working	Mandarin	Primary, secondary	Sometimes	Indifferent
MT2	26	No	Working	Shanghainese	Primary, secondary	Sometimes	Indifferent
MT3	17	No	Studying	Mandarin	Primary	Often	Amusing
MT4	26	Yes	Working	Shanghainese	Separated	Sometimes	Indifferent
FT1	15	No	Studying	Mandarin	Primary, secondary	Sometimes	Indifferent
FT2	21	No	Studying	Mandarin	Primary, secondary	Often	Amusing
FT3	17	No	Studying	Mandarin	Primary, secondary	Sometimes	Amusing
FT4	15	Yes	Studying	Mandarin	Separated	Often	Indifferent

separated twins, all the twins shared at least primary school education. Most of them had the same educational background before going to university. As a result, all these non-separated twins had many common friends. For the two separated pairs, they lived in different areas in Shanghai, and did not have any shared education. Thus, their language environments were more varied than the non-separated twins.

The twin speakers were also asked about their language use at home, at school or work, and what language they preferred while communicating with their twin siblings. The definition of language dominance is diverse because it is a measure of relative language performance and it can be evaluated in different ways (Daller, Yildiz, de Jong, Kan, & Basbagi, 2011; Dunn & Fox Tree, 2009). In our study, language dominance was judged based on the twins' preferred language as shown in our questionnaire. It was found that all the female twins and two pairs of male twins (MT1 and MT3) were Mandarin-dominant. They spoke Shanghainese only to some family members (mostly their grandparents who could not understand Mandarin), and communicated with other people including 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 brothers mostly in Shanghainese, while Mandarin was only used when they talked to people who could not understand Shanghainese.

All the twin speakers reported that they had experiences of being mistaken for their twin siblings by acquaintances, friends, relatives or even parents. Among all the twins, FT2 seemed to be the most similar-sounding pair – their mutual friends often had problems in telling them apart in face-to-face communications. Even their parents sometimes confuse their voices over the phone. Comparing with FT2, the other seven pairs of subjects had a lower chance to be mistaken – at least their parents seldom made mistakes in recognising their voices. FT4 and MT3 also claimed that they were often mistaken. However, their parents seldom confuse their voices. As for the attitudes towards being mistaken, FT2, FT3 and MT3 found it amusing that other people often got their identity wrong, and they wanted to be the same despite all the inconvenience. The other five pairs felt indifferent when being mistaken.

2.2. Materials

The diphthong /ua/ was used in the present study. This particular diphthong was chosen for two reasons: (1) It is present in both Shanghainese and Mandarin, and can appear in the same contexts (after [+back] consonants like /k, h/) in both languages, which makes the cross-language comparison possible. In fact, this is the only common diphthong which can appear in the same contexts in both languages for all speakers. (2) The two targets of this diphthong (i.e., /u/ and /a/) differ mainly in vowel height and lip-rounding, but not so much in backness. Therefore, the slope of formant transition would not be as sharp as other diphthongs like /aɪ/ (McDougall, 2004). It is of interest to us whether moderate movements in formant dynamics can also be used to discriminate speakers.

The experiments consisted of two wordlist reading tasks – one Mandarin task and one Shanghainese task, both containing target words and filler items (other real words in the two languages not having the diphthong /ua/). The target words (listed in Table 2) were three syllables that are phonemically the same in Shanghainese and in Mandarin. All the target words had a falling tone, a tone common to both languages. Token frequency was not controlled in this study because there is no data on the frequency of spoken Shanghainese. But all the words used here were common words in Mandarin and Shanghainese. The speakers had no difficulty in identifying them.

All the target words and filler items were embedded in carrier phrases to ensure the same intonation pattern, and they were presented to the speakers randomly. The two carrier phrases were of the same meaning: for Shanghainese, the carrier phrase was /ŋu doʔ____ gəʔ əʔ zɨ/ (我讀___個個字。 'I read the word ____.'); for Mandarin, the carrier phrase was /wo tu ____ tʂɤ kɤ tsɨ/ (我讀___這個字。 'I read the word ____.'). Using two different carrier phrases for the two languages is inevitable. The different segmental contexts in the two carrier phrases do not affect the validity of the comparison, because the data in the two languages were analysed separately. There was no cross-language comparison between twin speakers.

2.3. Procedures

The speakers 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). The two recording sessions were held at roughly the same time of the day. The subjects were asked to perform the same tasks in the two sessions, with the method and procedures being the same. The speakers did not report any ailments at the time when the recordings were made. Besides, previous studies show that long-term within-speaker differences could reduce the discriminatory power of formant frequencies (Rose & Clermont, 2001). Therefore, collecting data from multiple recording sessions would help to maximise the within-speaker differences. It is of interest to us whether the maximised within-speaker differences would still be smaller than the minimised between-speaker differences (i.e., the difference between identical twins).

Table 2
Target words used in the experiment.

	/kua/	/kʰua/	/hua/
Shanghainese	乖 'obedient'	誇 'praise'	歪 'slanting'
Mandarin	掛 'hang'	跨 'span'	畫 'draw'

All the recordings were made in a quiet room with a solid-state recorder with a sampling rate of 44,100 Hz. At the beginning of each session, the speakers were given some time to practice before the actual recording. When the recording began, the speakers were seated in front of a desk with the microphone placed about 20 cm from his or her mouth. They 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. For each speaker, 18 tokens were collected (3 syllables × 6 repetitions) for each language in each session. When the two sessions are collapsed, each speaker has 36 tokens (3 syllables × 6 repetitions × 2 sessions) for each language, which resulted in 1152 tokens in total (3 syllables × 6 repetitions × 2 sessions × 16 speakers × 2 languages).

We did not control for the Language Mode effects (Grosjean, 2001) as we collected both Shanghainese and Mandarin data in the same recording section. Nevertheless, the experimenter communicated with the twin speakers in the respective language when instructing them what to do for the Shanghainese and Mandarin parts.

2.4. Measurements

The recordings were downsampled to 22,050 Hz and analysed using Praat. The diphthong /ua/ were marked manually for each sample. The start and the end of periodicity were regarded as the beginning and the end of the vowel. The total duration of the vowel was divided into 10 equal parts and $F1$–$F4$ frequencies were tracked at each +10% step (11 points in total) using a Praat script. The results were all checked manually using FFT spectral slices, and any anomaly was corrected.

Two types of statistical analyses were conducted: MANOVAs using formant frequencies at individual measuring points (Section 3.1) and Discriminant Analysis using formant frequencies at 15 measuring points (Section 3.2). Details of these analyses will be given in Section 3 for ease of reference.

3. Results

3.1. Formant trajectories

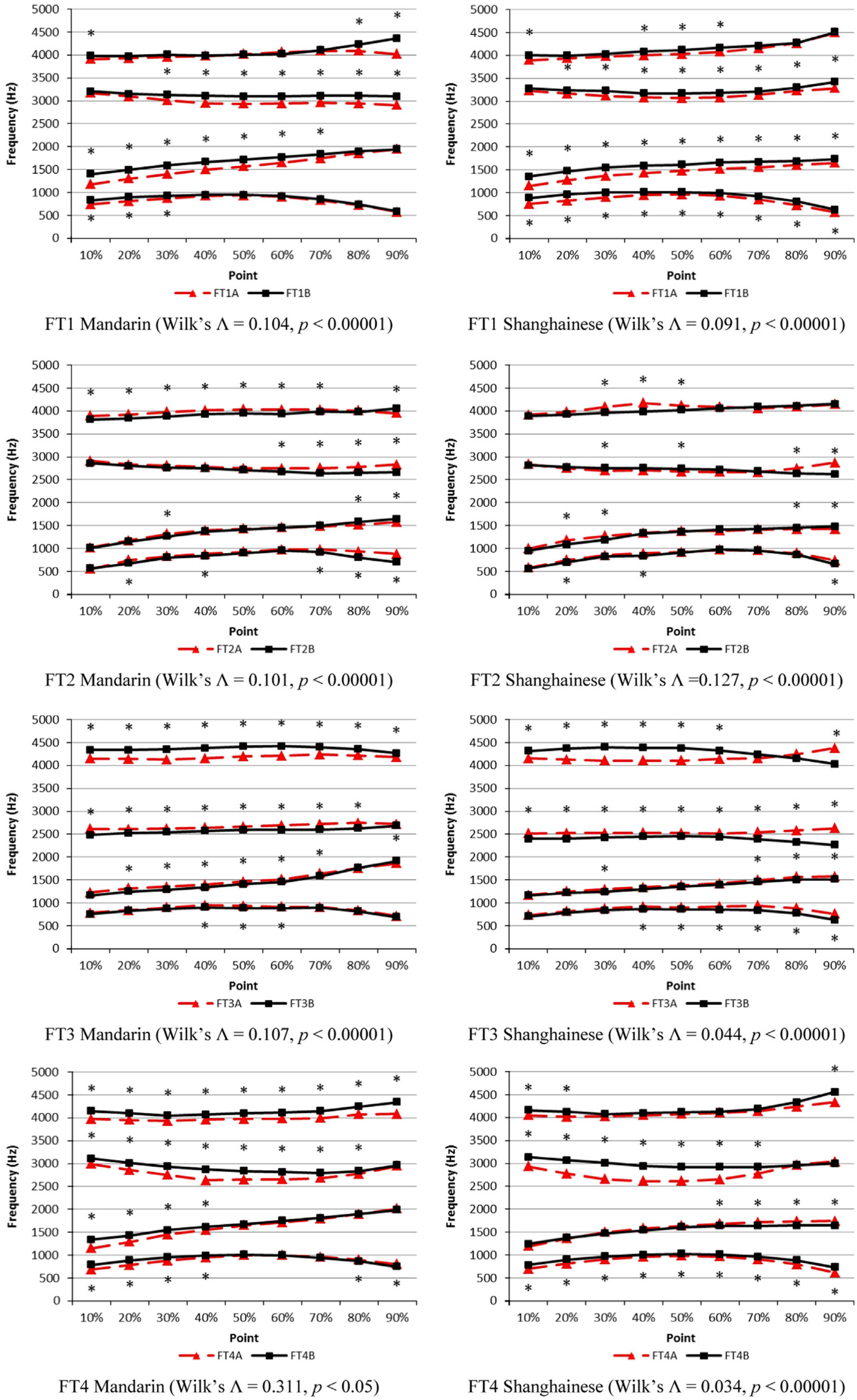
Figs. 1 (female) and 2 (male) show the averaged trajectories of the first four formants of all the identical twins. MANOVAs were conducted for each pair of twins for each language. Formant frequencies of all but the first and the last measuring points (i.e., 10–90% points) were included as dependent variables, to minimize the perturbation from flanking consonants. For each comparison, 36 tokens per speaker (3 syllables × 6 repetitions × 2 sessions) were compared. The MANOVAs results showed that all the twins showed significant differences in their formant dynamics (Wilk's lambdas and the p values are given under the respective panels in Figs. 1 and 2. Full details can be found in supplementary Appendix A). The asterisks in Figs. 1 and 2 stand for significant differences between the two twin members on the specific points from tests of between-subject effects. For each measuring point, only one comparison was conducted between the twin siblings, and no cross-point multiple comparison was conducted, so no adjustment was applied.

The results suggested that despite identical anatomy and similar language environments, the twins did exhibit differences in their formant dynamics. All the MANOVAs reported significant differences between identical twins. One interesting finding is that, by visual inspection, non-separated pairs are not necessarily more similar than separated pairs. Take the separated FT4 as an example, their Mandarin /ua/ showed significant differences on 27 out of 36 points, and their Shanghainese /ua/ showed differences on 23 out of 36 points on the first four formants. These numbers were comparable to the results of FT1 and FT3 who were non-separated. Separated MT4 showed a similar tendency. If we compare MT4 Shanghainese with MT2 Shanghainese (both pairs are Shanghainese-dominant), we can see that MT4 had fewer within-pair differences. In other words, separated twins can be as similar as, or even more similar than, non-separated twins in their formant trajectories.

3.2. Discriminant analysis

Although all the twins exhibited differences in their first four formants, some twins (e.g. MT3) were found not to differ much in individual measuring points. Therefore, Discriminant Analysis was carried out by incorporating data from multiple measuring points into the same analysis to see whether the twins could be discriminated by their dynamic features. Discriminant Analysis was chosen for three reasons. First, it is a multivariate test which can make use of the dynamic features in formant trajectories. Second, it takes both between-speaker and within-speaker variations into consideration. Third, it contains a classification process which predicts the most probable group membership for each case, and the correct classification rates can be taken as a measure to quantify the similarity between the twins. In the current study, each subject was treated as a group of many tokens, with ‘speaker’ as the dependent variable. The formant frequency values on multiple points were used as ‘predictors’. Statistical analysis derived linear discriminant functions that could best separate different groups based on the predictors. The discriminant functions were then used to predict the most probable group membership for each token. The more cases being correctly classified, the better the speaker groups were separated, i.e., more between-speaker differences exist for the twin siblings.

Discriminant Analysis was conducted separately for each language and sex, resulting in four conditions – female Mandarin, female Shanghainese, male Mandarin, and male Shanghainese. Since Discriminant Analysis is highly sensitive to outliers, the data was scanned before the test was conducted. Both univariate outliers (z-score > 3.29) and multivariate outliers ($\chi^2 \geq \chi^2_{crit}$, $p < 0.001$) were removed.



FT1 Mandarin (Wilk's Λ = 0.104, $p < 0.00001$)

FT1 Shanghainese (Wilk's Λ = 0.091, $p < 0.00001$)

FT2 Mandarin (Wilk's Λ = 0.101, $p < 0.00001$)

FT2 Shanghainese (Wilk's Λ =0.127, $p < 0.00001$)

FT3 Mandarin (Wilk's Λ = 0.107, $p < 0.00001$)

FT3 Shanghainese (Wilk's Λ = 0.044, $p < 0.00001$)

FT4 Mandarin (Wilk's Λ = 0.311, $p < 0.05$)

FT4 Shanghainese (Wilk's Λ = 0.034, $p < 0.00001$)

Fig. 1. Formant trajectories of the female twins. The asterisks indicate significant differences.

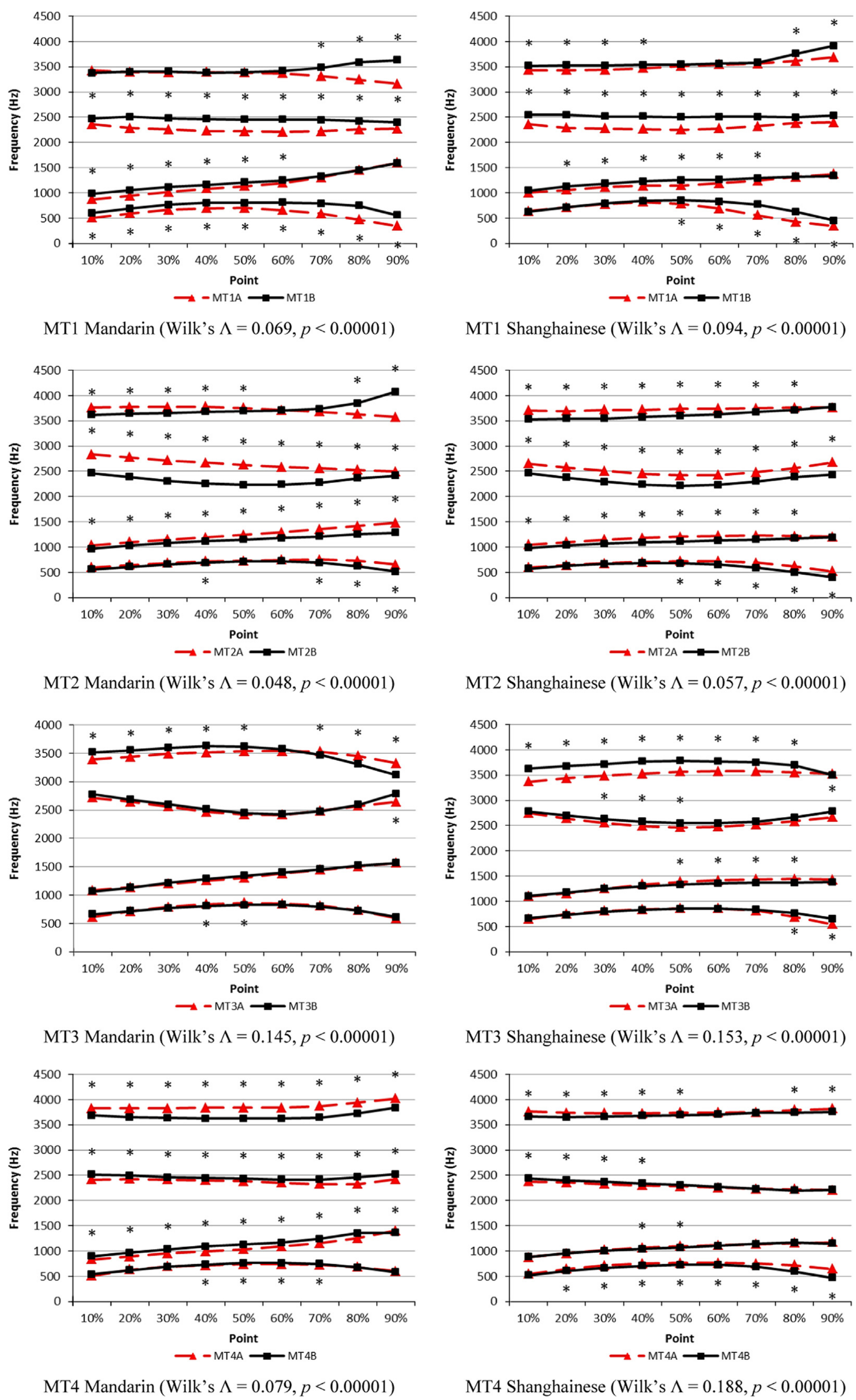


Fig. 2. Formant trajectories of the male twins. The asterisks indicate significant differences.

The 10%, 30%, 50%, 70%, and 90% points on $F1$, $F2$ and $F3$ were used as predictors, so there are 15 predictors altogether which track the whole formant contour. $F4$ was not included for two reasons. First and foremost, our sample size only allows us to use 15 predictors at most, as Discriminant Analysis requires the size of the smallest group to be larger than the number of predictors. In our study, the smallest group size used in calculating the discriminant functions was 16. In addition, 15 predictors can adequately capture the formant dynamics already. Second, given only 15 predictors were allowed, we gave up $F4$ because the extracted frequency values of $F4$ might not be as reliable as those of $F1$–$F3$ due to weak energy of the upper formants.

The Discriminant Analysis results showed that the overall correct classification rates were very high when multiple points on $F1$–$F3$ were combined (see Fig. 3). For the female Mandarin, male Mandarin and male Shanghainese conditions, when 15 predictors were included, the classification rates were 79.4%, 78.2% and 80.9% respectively, and for the female Shanghainese condition, the correct classification rate reached 89.5%. These numbers were significantly higher than chance (i.e., 100%÷8 speakers = 12.5%).

Besides using absolute formant frequency values, we also followed McDougall (2006) in characterising the first three formant contours with cubic polynomial equations (i.e., each formant was fitted with an equation $y = a_0 + a_1t + a_2t^2 + a_3t^3$, in which t stands for the time point). The coefficients (i.e., a_0, a_1, a_2, and a_3 of the three equations, 12 predictors in all) were used as predictors to conduct classifications. Table 3 shows the correct classification rates in Discriminant Analysis using cubic polynomial coefficients. The overall classification rates using absolute formant frequencies as predictors are also given for reference. Both the classification rates and patterns of the two methods are very similar. In fact, all the conditions yielded slightly lower classification rates when coefficients were used. The results may be accounted for by the fact that the cubic polynomial equations are approximations of the actual formant dynamics. Although the R values (i.e., how good the dynamics are fitted with the equations) were quite high for each trajectory (all over 0.8), some subtle information was filtered out. Such information may be important in identifying identical twins who can be very similar in their acoustic features. Given the fact that both polynomial coefficients and absolute formant frequency gave highly comparable results, and that polynomial coefficients are only approximants of the actual dynamics, we decided to use the absolute values instead of the polynomial equations in our analysis.

Although the overall classification rates showed that Discriminant Analysis was very successful in discriminating the speakers, misclassifications also existed. Most of the misclassifications involved identical twins (i.e., identifying tokens from one member as tokens from the other member of the twins). Table 4 gives the rate of each speaker being misidentified as their twin sibling. In the

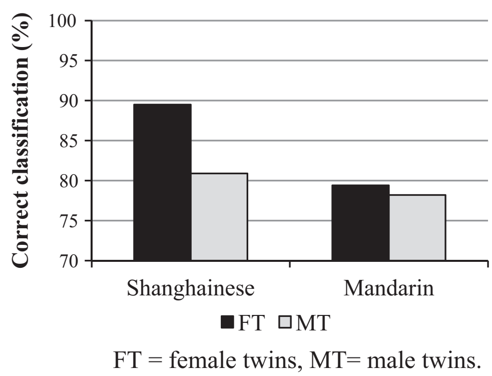


Fig. 3. Overall correct classification rates in Discriminant Analysis. FT = female twins, MT = male twins.

Table 3
Correct classification rates in Discriminant Analysis using cubic polynomial coefficients of the first three formants and using absolute values of formant frequencies.

	Polynomial coefficients (%)	Absolute formant frequencies values (%)
FT Mandarin	79.2	79.4
FT Shanghainese	88.4	89.5
MT Mandarin	78.1	78.2
MT Shanghainese	80.4	80.9

Table 4
Rate of each speaker being misidentified as their twin sibling in Discriminant Analysis. The shaded cells indicate the twins' dominant language.

Twin pair	Female Mandarin	Female Shanghainese	Male Mandarin	Male Shanghainese
1A	13.9	5.7	8.8	2.9
1B	19.4	16.1	11.1	5.6
2A	14.3	16.7	2.8	11.1
2B	8.3	11.4	0.0	2.8
3A	14.3	0.0	34.4	24.2
3B	17.6	5.6	26.5	17.1
4A	27.3	0.0	8.3	16.7
4B	22.2	2.8	8.3	16.7

current study, we regard the misidentification rates as the degree of similarity between two speakers. The higher the number is, the more the speaker resembled his/her twin sibling.

The results echo the patterns in Section 3.1. When we compared the Mandarin results with the Shanghainese results, large discrepancies were found between the misidentification rates in the two languages. The most extreme case was FT4 who were Mandarin-dominant: around 25% of their Mandarin productions were misclassified as their twin sisters', but almost none of their Shanghainese productions were misclassified. All the twins except FT2 showed higher chances of being misclassified in their dominant language. This suggests that the twin speakers were more similar in their dominant language.

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 twins appeared to be more similar than some of the non-separated twins in the current study. The misclassification rates in Table 4 suggested that FT4 (separated) were more similar than all the non-separated female twins in Mandarin, and MT4 (separated) were more similar than the other Shanghainese-dominant pair MT2 (non-separated) in both languages. Comparing with MT4, the misidentifying rates of MT2 were lower, especially for MT2B whose Shanghainese were misidentified as MT2A's in only 2.8% of cases. This finding indicates that given enough interaction between the twins, separated twins can be more similar in formant dynamics than non-separated twins.

4. Discussion

This study investigated the formant dynamics in the speech of bilingual identical twins. The results showed that although being very similar, identical twins had significant differences in their formant dynamics in both languages, and the differences were large enough to distinguish them. Discriminant Analysis yielded high correct classification rates when multiple points on the first three formants were combined. The misclassification rates were used to quantify the similarity between twin speakers. It was found that the twins had different variability in the two languages they speak. They were more similar in their dominant language. The Mandarin-dominant twins were more often misidentified as their twin siblings in Mandarin despite the fact that they had a more uniform Shanghainese environment (they only spoke Shanghainese to their elderly family members who could not understand Mandarin, while they used Mandarin in all other occasions).

A possible reason for the higher similarity in their dominant language may be that, as the twins communicated with each other mostly in their dominant language, frequent daily conversations provided them with more chances to adapt to each other. Adaptation means that speakers modify their speech patterns to make themselves sound similar to the interlocutors. As for the non-dominant language, since they seldom talked to each other using that language, there was little chance for adaptation no matter how uniform the language environment was. In fact, this account is also compatible with FT2's pattern. While other Mandarin-dominant pairs claimed that they seldom communicated 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, their misidentification rates were also comparable across languages (see Table 4). Besides, language dominance can also explain why the highest classification rates were found in the female Shanghainese condition (Fig. 3): none of the female twins were Shanghainese-dominant, and therefore, the between-speaker differences for the twins were relatively larger in their Shanghainese /ua/. By contrast, half of the male twins were Mandarin-dominant and the other half were Shanghainese-dominant. Therefore, the overall classification rates for the two male conditions were quite comparable.

In addition, the results indicated that the two pairs of twins who were separated from birth were found to be as similar as, or even more similar than some of the non-separated twins in their formant dynamics, despite having different language environments. These findings underline the importance of a third factor besides anatomy and language environment in shaping one's speech patterns – individual choices (Nolan, 1997) as discussed in the Introduction. These choices can be conscious or subconscious. In the present study, the attitude towards being identical twins appears to be a factor affecting speakers' choices. The fact that the most similar pair in the present study (i.e., MT3, see Table 4) claimed that they wanted to sound alike in the background questionnaire indicates that the similarity between identical twins may be partially determined by their attitudes towards being twins. In fact, the good relationship between the separated twins FT4 observed by the first author also lends support to this possibility. They claimed that they liked to stay with each other. They introduced their own friends to each other, and also participated in each other's social events. All these showed that they liked to be twins. And this can be a potential reason for the highly convergent formant patterns in their speech.

The effect of social factors on one's speech patterns has been confirmed by studies on speech adaptation (also termed as 'accommodation', 'convergence', or 'alignment'). Previously, studies have shown that adaptations can be motivated by social attitudes (Babel, 2009; Giles & Coupland, 1991; Shepard, Giles, & LePoire, 2001). Similarly, for those identical twins who wished to signify their identity as twins, it is likely that their speech patterns would converge to a higher degree than those who were ambivalent about it. And this might be the most important influence that "individual choice" may have on the similarity of speech patterns between identical twins. Therefore, it is possible that the uniformity of language environment is not as important as individual choices in shaping the speech patterns of identical twins. Identical twins can choose to be the same or different despite their language environment.

Our findings have some implications on issues related to speech production. No current model of speech production focuses on the speech of identical twins. Perhaps the models that can best account for the variations observed in the speech of identical twins in the present study are the exemplar models (e.g. Johnson, 1997; Pierrehumbert, 2001). There are different versions of exemplar models which differ in details (e.g. the basic speech units that are stored as exemplars). However, they share some main ideas that are relevant to the current study. The exemplar models hypothesise that the mental representations of sound categories are exemplar clouds containing large numbers of tokens. All the tokens that the listeners hear are stored as exemplars in the memory with phonetic details retained. The exemplars are arranged by their acoustic similarities so that similar tokens are close to each other. Different exemplars have different levels of activation which is a function of the frequency and the recency of that particular exemplar. The more frequently and the more recently the token is encountered, the higher the activation.

Although the exemplar models are developed mainly for speech perception, the basic ideas can be applied in speech production as well. Pierrehumbert (2001) assumed that speech production proceeds in the opposite direction to speech perception – in production, the speaker selects an exemplar randomly from the exemplar clouds, and then makes the articulatory plan. The selection is partially constrained by the activation of the exemplars. That is to say, such models are very sensitive to usage frequency. This frequency-sensitive feature can be used to account for the higher similarity between the twins in their dominant language than in their non-dominant language. According to Pierrehumbert's (2001) exemplar model, the speaker selects exemplars based on their activation levels. Given the fact that the activation level is positively correlated with the frequency of the exemplar, we expect that the exemplars that are more frequently encountered would have higher probabilities to be selected. Most of the identical twins in the current study claimed that they seldom communicated with each other in their non-dominant language. The exemplars of their twin sibling's speech in their non-dominant language were of low frequency, and therefore, low activation levels. By contrast, since the twins communicate with each other very often in their dominant language, the exemplars of their twin sibling's speech in their dominant language were of high frequency, and high activation levels. As a result, they were more likely to select the exemplars of their twin sibling's speech as the phonetic targets in their dominant language than in their non-dominant language, which caused the higher similarity in their dominant language.

Besides usage frequency, social factors (e.g. social identity, social attitudes towards certain speakers) may also affect speech production by biasing the selection of exemplars (Pierrehumbert, 2001). As discussed above, the attitude towards being identical twins may have affected how they adapt to each other's speech patterns. For those who wish to sound alike, whether they are separated or not, the exemplars of their twin sibling's speech are more likely to be selected as the phonetic targets. This can be a potential reason for the highly convergent formant patterns in their production.

To summarise, the current study shows that individual choices may account for the differences in the formant dynamics of identical twins. However, it also has some limitations. First, this study had only eight pairs of identical twins, although it compares well with previous studies in terms of sample size. All the twin speakers were young adults, but the male twins were slightly older than the female twins, and there was no Shanghainese-dominant female twin. The small sample size does not allow us to investigate the separate effects of sex and age reliably. These identical twins speakers with native proficiency in both languages were very difficult to find, as there is no twin registry in Shanghai. We could only rely on word of mouth to recruit them. Larger-scale studies on bilingual identical twins are necessary in order to confirm our findings, and to investigate the effects of sex and age.

Second, the data were based on only one common diphthong /ua/ in both Shanghainese and Mandarin. Future studies can either use a different language pair with more common diphthongs. Finally, more precise measures are needed to elicit speakers’ attitude towards being twins, so that quantitative analysis can be conducted to confirm our suggestion that their attitudes can affect the similarities in formant dynamics between them. More concrete results can be obtained to illustrate how individual choices can affect speech production. It is hoped that these results can corroborate our suggestion of the importance of “choices” in speech production.

Acknowledgements

The authors would like to thank the twin speakers for their participation in the project. We also thank the phonetics lab of the Shanghai Normal University for allowing us to do some recording in their lab. This project was supported by the Department of Linguistics and Modern Languages at the Chinese University of Hong Kong. We thank the anonymous reviewers for their helpful suggestions.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.wocn.2015.03.003.

References

Babel, M. (2009). *Phonetic and social selectivity in speech accommodation* Unpublished Ph.D. thesis. Berkeley: Department of Linguistics, University of California.
Beck, J. M. (1997). Organic variation of the vocal apparatus. In W. J. Hardcastle, & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 256–297). Oxford: Blackwell.
Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, *97*, 1916–1924.
Brown, R. (1987). *Auditory speaker recognition*. Hamburg: Buske.
Chen, Y. (2008). The acoustic realization of vowels of Shanghai Chinese. *Journal of Phonetics*, *36*, 629–648.
Daller, M., Yildiz, C., de Jong, N., Kan, S., & Basbagi, R. (2011). Language dominance in Turkish–German bilinguals: Methodological aspects of measurements in structurally different languages. *International Journal of Bilingualism*, *15*, 215–236.
Debruyne, F., Decoster, W., Van Gijsel, A., & Vercammen, J. (2002). Speaking fundamental frequency in monozygotic and dizygotic twins. *Journal of Voice*, *16*, 466–471.
Dunn, A., & Fox Tree, J. E. (2009). A quick, gradient Bilingual Dominance Scale. *Bilingualism: Language and Cognition*, *12*, 273–289.
Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S., & Hallé, P. (2008). Cross language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics*, *36*, 649–663.

Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America*, *126*, 2603–2618.
Garvin, P. L., & Ladefoged, P. (1963). Speaker identification and message identification in speech recognition. *Phonetica*, *9*, 193–199.
Giles, H., & Coupland, N. (1991). *Language: Contexts and consequences*. Keynes: Open University Press.
Glenn, J. W., & Kleiner, N. (1968). Speaker identification based on nasal phonation. *Journal of the Acoustical Society of America*, *43*, 368–372.
Greisbach, R., Esser, O., & Weinstock, C. (1995). Speaker identification by formant contours. In A. Braun, & J.-P. Köster (Eds.), *Studies in forensic phonetics: BEIPHOL*, 64 (pp. 49–55). Trier: Wissenschaftlicher Verlag Trier.
Grosjean, F. (2001). The bilingual's language modes. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1–22). Oxford: Blackwell.
Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals. Age of acquisition effects on the mutual influence of the first and second languages. *Phonetica*, *60*, 98–128.
Ingram, J. C. L., Prandolini, R., & Ong, S. (1996). Formant trajectories as indices of phonetic variation for speaker identification. *Forensic Linguistics*, *3*, 129–145.
Jessen, M. (1997). Speaker-specific information in voice quality parameters. *Forensic Linguistics*, *4*, 84–103.
Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego: Academic Press.
Kinoshita, Y. (2001). *Testing realistic forensic speaker identification in Japanese: A likelihood ratio-based approach using formants* Unpublished Ph.D. thesis. Australian National University.
Kühnert, B., & Nolan, F. (1999). The origin of coarticulation. In W. J. Hardcastle, & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques in speech production* (pp. 7–30). Cambridge: Cambridge University Press.
Lee, W. S., & Zee, E. (2003). Illustrations of the IPA: Standard Chinese (Beijing). *Journal of the International Phonetic Association*, *33*, 109–112.
Loakes, D. (2006). *A forensic phonetic investigation into the speech patterns of identical and non-identical twins* Unpublished Ph.D. thesis. School of Languages and Linguistics, The University of Melbourne.
Loakes, D. (2008). A forensic phonetic investigation into the speech patterns of identical and non-identical twins. *International Journal of Speech, Language and the Law*, *15*, 97–100.
Lundström, A. (1948). *Tooth size and occlusion in twins*. Basle: S. Karger.
Macleod, A. A. N., & Stoel-Gammon, C. (2005). Are bilinguals different? What VOT tells us about simultaneous bilinguals. *Journal of Multilingual Communication Disorders*, *3*, 118–127.
McDougall, K. (2004). Speaker-specific formant dynamics: An experiment on Australian English /ai/. *International Journal of Speech, Language and the Law*, *11*, 103–130.
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*, 89–126.
Mayr, R., Price, S., & Mennen, I. (2012). First language attrition in the speech of Dutch-English bilinguals: The case of monozygotic twin sisters. *Bilingualism: Language and Cognition*, *15* (4), 687–700.
Morrison, G. S. (2009). Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs. *Journal of the Acoustical Society of America*, *125*, 2387–2397.
Morrison, G. S., & Assmann, P. F. (Eds.). (2013). *Vowel inherent spectral change*. New York: Springer-Verlag.
Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, *80*, 1297–1308.
Nolan, F. (1997). Speaker recognition and forensic phonetics. In W. J. Hardcastle, & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 744–767). Oxford: Blackwell.
Nolan, F., & Oh, T. (1996). Identical twins, different voices. *Forensic Linguistics*, *3*, 39–49.
Nolan, F., Oh, T., McDougall, K., De Jong, G., & Hudson, T. (2006). A forensic phonetic study of "dynamic" sources of variability in speech: The DyViS porject. In *Proceedings of the 11th Australian international conference on speech science and technology* (pp. 13–18). Auckland.
Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.
Przybyla, B. D., Horii, J., & Crawford, M. H. (1992). Vocal fundamental frequency in a twin sample: Looking for a genetic effect. *Journal of Voice*, *6*, 261–266.
Rose, P. (1999). Differences and distinguishability in the acoustic characteristics of hello in voices of similar-sounding speakers: A forensic phonetic investigation. *Australian Review of Applied Linguistics*, *22*, 1–42.
Rose, P. (2002). *Forensic speaker identification*. London: Taylor & Francis.
Rose, P., & Clermont, F. (2001). Comparison of two acoustic methods of forensic speaker identification. *Acoustics Australia*, *29*, 31–35.
Rose, P., Kinoshita, Y., & Alderman, T. (2006). Realistic extrinsic forensic speaker discrimination with the diphthong /ai/. In *Proceedings of the 11th Australasian international conference on speech science & technology* (pp. 329–334). Auckland.
Ryalls, J., Shaw, H., & Simon, M. (2004). Voice onset time production in older and younger female monozygotic twins. *Folia Phoniatrica et Logopaedica*, *56*, 165–169.
Shepard, C. A., Giles, H., & LePoire, B. A. (2001). Communication accommodation theory. In W. P. Robinson, & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 33–56). Chichester: John Wiley & Sons.
Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, *9*, 97–114.
Thompson, P. M., Cannon, T. D., Narr, K. L., van Erp, T., Poutanen, V., Huttunen, M., et al. (2001). Genetic influences on brain structure. *Nature Neuroscience*, *4*, 1253–1258.
Townsend, G. C., Richards, L., Hughes, T., Pinkerton, S., & Schwerdt, W. (2003). The value of twins in dental research. *Australian Dental Journal*, *48*, 82–88.
Van Gysel, A., Vercammen, J., & Debruyne, F. (2001). Voice similarity in identical twins. *Acta oto-rhino-laryngologica Belgica*, *55*, 49–55.
Van Lierde, K. M., Vinck, B., De Ley, S., Clement, G., & Van Cauwenberge, P. (2005). Genetics of vocal quality characteristics in monozygotic twins: A multiparameter approach. *Journal of Voice*, *19*, 511–518.
Weirich, M. (2011). *The influence of NATURE and NURTURE on speaker-specific parameters in twins speech: Articulation, acoustics and perception* Unpublished Ph.D. thesis. Humboldt-Universität zu Berlin.
Weirich, M. (2012). The influence of NATURE and NURTURE on speaker-specific parameters in twins' speech: Acoustics, articulation and perception. *International Journal of Speech, Language and the Law*, *19*, 119–122.
Whiteside, S. P., & Rixon, E. (2001). Speech patterns of monozygotic twins: An acoustic case study of monosyllabic words. *The Phonetician*, *82*, 9–22.
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.
Whitworth, N. (2000). *Acquisition of VOT, and vowel length by English-German bilinguals: A pilot study*. Leeds working papers in linguistics and phonetics (Vol. 8, pp. 1–16).
Wolf, J. J. (1972). Efficient acoustic parameters for speaker recognition. *Journal of the Acoustical Society of America*, *51*, 2044–2056.
Xu, B.H, & Tang, Z. Z. (1988). *A grammar of urban Shanghainese (in Chinese)*. Shanghai: Shanghai Education Publisher.
Zhang, C., Su, B., Cao, Q., & Zhao, X. (2010). An investigation on the role of formant dynamics in speaker identification (in Chinese). *Paper presented in the 9th phonetic conference of China* (PCC 2010). Tianjin.
Zhang, C., Morrison, G.S., & Thiruvaran, T. (2011). Forensic voice comparison using Chinese /iau/. In *Proceedings of the 17th international congress of phonetic sciences* (*ICPhS*) (pp. 2280–2283). Hong Kong.
Zhu, X. (1999). *Shanghai tonetics*. Munchen: Lincom Europa.