Social Sciences & Humanities Open 10 (2024) 101062

Contents lists available at ScienceDirect

Social Sciences & Humanities Open

journal homepage: www.sciencedirect.com/journal/social-sciences-and-humanities-open

Regular Article

Monozygotic twin sisters differ in voice quality: Acoustic analysis of the diphthong [ʻoj] produced by Brazilian native speakers

Paula Bruna Campos [a,e,f,g], Luchesi Lilian Cristina [a,b,c,d], Lucci Tania Kiehl [b,c,d], Otta Emma [b,c,d], Monticelli Patrícia Ferreira [a,*]

[a] Department of Psychology, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Laboratório de Etologia e Bioacústica (EBAC), Avenida dos Bandeirantes, 3900, 14040-901, Ribeirão Preto, SP, Brazil
[b] Department of Experimental Psychology, Instituto de Psicologia, Universidade de São Paulo, São Paulo, SP, Brazil
[c] Psychoethology - Human Ethology Laboratory, São Paulo, SP, Brazil
[d] Av. Prof. Mello Moraes, 1721, 05508-030, Cidade Universitária, São Paulo, SP, Brazil
[e] Department of Mechanical Engineering, São Paulo, SP, Brazil
[f] Escola Politécnica, Universidade de São Paulo, São Paulo, SP, Brazil
[g] Av. Professor Mello Moraes, 2231, 05508-030, Cidade Universitária, São Paulo, SP, Brazil

ARTICLE INFO

Keywords:
Acoustic communication
Human behavior
Bioacoustics
Vocal production
Voice recognition
Zygosity

ABSTRACT

Identical twins challenge individual voice recognition applied in communication, criminal investigation, and health technologies. The similarities between siblings are due to the genetic heritage that produced either their vocal tract or the psychological processes acting since birth. To contribute to a precise individual identification by voice, we tested the hypothesis that even MZ siblings would differ in the spectral distribution of formants and sound quality. We recorded 86 same-sex twin pairs (71 MZ, 15 DZ 30 ± 12 years old) saying the Portuguese verse of "Hi, my name is (Pedro or Ana)." We trimmed the diphthong Ój ("Hi") and measured f0, F1-F4 formants, jitter, shimmer, and Harmonic-to-noise ratio (HNR). As the male sampling was small, we compared only female twins using the Wilcoxon-matched test, the intrapair Correlation Coefficient (ICC), and GLMM models. On average, [Ój] lasted 0.24 ± 0.07 s, and f0 ranged from 469 to 510 kHz according to sex. We found no differences between the MZ or DZ sisters in any acoustic parameter, besides F1 for MZ ones. Also, MZ sisters were more alike (F1, jitter, HNR, weak; F3, F4 moderate) than DZ (only F1 high ICC). The interaction zygosity × age affected intrapair similarity, the older the sisters, the greater their differences. Short voice segments, like the diphthong [Ój], may not be good for differentiating identical twins but provide insights to discussion; F2, F4, and shimmer should be investigated in future studies considering males, younger participants, and other vocal expressions.

1. Introduction

Voice plays an important part in human communication means. The sounds used in human verbal communication result from combinatorial possibilities of the dynamic vocal tract elements (Barbosa & Madureira, 2015). Among them, two anatomic structures act in vocal production: the position and the degree of contraction of folds and sub-laryngeal elements (the source part of the system in the Source-filter theory of Gunnar Fant[1]; Tokuda, 2021) the sub-larynx structures and the position and shape of the tongue (the filter part) (Barbosa & Madureira, 2015). This dynamic system offers an opportunity for the development and evolution of individuality in voice, once different people can sound different when speaking even a short vowel (see Zhang, 2016).

In Brazil, the acoustic parameters used in clinical voice evaluation by phonologists, besides f0, are jitter (%), shimmer (dB), and signal-to-noise ratio (HNR) (de Felippe et al., 2006). The first two are also used indirectly to evaluate laryngeal conditions during the emission of sustained vowels in Europe due to their relation with vocal fold irregular

* Corresponding author.
E-mail addresses: brunacampospaula@gmail.com (P. Bruna Campos), lili.luchesi@gmail.com (L. Lilian Cristina), tanialucci@usp.br (L. Tania Kiehl), emmaotta@usp.br (O. Emma), monticel@usp.br (M. Patrícia Ferreira).

[1] Fant, Carl Gunnar Michael. Retrived from: https://scholar.google.com/citations?user=uQ1eyqgAAAAJ&hl=pt-BR&oi=sra Fant, G., 1960 (Acoustic Theory of Speech Production, The Hague: Mouton) is a landmark publication cited in many acoustic phonetics papers and books for laying out the mathematical derivation of the acoustic properties of most speech sounds.

https://doi.org/10.1016/j.ssaho.2024.101062
Received 6 November 2023; Received in revised form 18 July 2024; Accepted 31 July 2024
Available online 12 August 2024

vibration and dysphonia (Brockmann et al., 2008). The first two are frequency and amplitude variation measures; the primmer is perceived as the voice's roughness and the latter as hoarseness (Barbosa & Madureira, 2015). The vowel's HNR results from the interaction between the acoustic wave and the noise that arises as a subproduct of the folds and the vocal tract activities. Again, age and habits acquired in adulthood may change voice in these parameters. HNR was found to be higher in females and influenced by age (lower in women from 70 to 90 than from 21 to 34 years old; de Felippe et al., 2006); smokers had lower f0 and higher jitter and shimmer compared to non-smokers (Tajada et al., 1999).

Along the aging, behavioral habits like smoking or singing practice were found to affect vocal tract elements or stiff the laryngeal tube (Behlau, 2001; Stemple et al., 2018), heightening the differences that can be quantified in the acoustic analysis. Longitudinal approaches have found that f0 (fundamental frequency) and F1 (first formant) decrease with age in women, and the opposite tends to occur with f0 in men (Eichhorn et al., 2017; Kent & Vorperian, 2018).

The existence of individual signatures is in the best interest of health science and technology in artificial intelligence (Casanova et al., 2021; Possani, 2021); for instance, Siri, the intelligent assistant of Apple Inc., can recognize speakers besides words (Casanova, 2021). Linear Model parameters based on LPC analysis have been widely used in these systems and speech recognition (Moondra & Chahal, 2023; Alkhatib et al., 2023) making them a potentially relevant approach to be used in voice investigations.

1.1. Twins studies

Literature evidence points that identical twins are more similar than fraternal twins and non-twins in f0 mean and variability (Debruyne et al., 2002), voice spectrum above F1 (van Lierde et al., 2005), and are consistently more difficult to distinguish by automatic speaker recognition systems (Sabatier et al., 2019). Hence, twin studies are opportunities for the comprehension of human aging and development processes influenced by multiple factors.

Monozygotic twins share articulatory patterns and develop similarly in speech and language abilities and are consequently used to share speech and language disorders (Whiteside & Rixon, 2003). Nevertheless, even their voices' f0, supposedly equal in identical twins due to its strong relationship with anatomical traits (phonatory tube's length and mass; Barbosa & Madureira, 2015), can accumulate differences along the lifetime (Fuchs et al., 2000; Van Gysel et al., 2001; Whiteside & Rixon, 2013). Therefore individual differentiation in voice at the zygosity level challenges forensic fostering suspect identification (Cavalcanti et al., 2021; San Segundo et al., 2017) due both to the genetic heritage of the vocal tract anatomy and the psychological processes acting in voice development (Debruyne et al., 2002; Van Lierde et al., 2005; Weirich & Lancia, 2011). Once these phenotypic variances can be associated with genetic and environmental variances, twins are an interesting model for assessing variations due to genetic factors (Smith, 1970). The issue of heritability is a widely debated topic in twin studies (Matthews & Turkheimer, 2022), and while the literature has made some advancements in voice studies involving twins, these studies are predominantly conducted within Western, Educated, Industrialized, Rich, and Democratic populations (WEIRD).

Heritability is a statistical measure of a given population and environment that is influenced by population characteristics. Therefore, changing the population or the environment can alter heritability. For example, in a study conducted with children aged 6–12 years, Turkheimer et al. (2003) found that socioeconomic status modifies the heritability of IQ. Among children raised in low socioeconomic households, 60% of the variation in IQ was attributed to the shared environment, with little contribution from genes, whereas among children raised in high socioeconomic households, the opposite result was observed. To better comprehend the concept of heritability, it's important to have studies with twins from non-WEIRD populations (Hagenbeek et al., 2022).

Thus, our contribution in this present study of twin voices is to identify the acoustic parameters that are most susceptible to variations despite zygosity focusing on a very short sound comparable to vocalizations of non-human mammals (Taylor & Reby, 2010), a diphthong. Adopting a more representative sample of individuals along the adulthood ages we also contribute to investigations of the age effect on voice production. Moreover, unlike previous studies, which were based on one or up to 10 twin pairs, and few of them investigated Brazilian speakers (Cavalcanti et al., 2021), ours included a more diverse cohort of participants. Finally, the distinction of MZ twins by acoustic analysis and in contrast with DZ twins may be a cheaper and noninvasive way of determining zygosity when DNA analyses are unfeasible (Ariyaeeinia et al., 2008), and may promote advances in the comprehension of articulation disorders causes (Fernandes et al., 1999).

Here we compared monozygotic (MZ) and dizygotic (DZ) twin siblings' diphthong emissions and measured the probability of finding the correct pair of a twin unit based on the Euclidean distance metric.

2. Material and methods

2.1. Participants

We analyzed MZ and DZ recordings of a diphthong production. We adopted a collection of twin recordings of the *Painel USP de Gêmeos* (Twins USP Panel Project, Otta et al., 2019) at the University of São Paulo. The Panel is a huge registry of Brazilian twins that contributes to the understanding of human behavior and the peculiarities of gemelarity.

The study included the voices of 86 same-sex, reared-together twin pairs (172 individuals) who volunteered for the research (Table 1). The sample comprised 71 MZ pairs (50 female pairs) and 15 DZ pairs (12 female pairs), ranging from 18 to 66 years (mean age: 31.7 ± 11.6 years). Zygosity was determined through a DNA test performed on blood samples collected on the same day as the voice recordings.

All participants reported having lived most of their lives in the São Paulo state, except for two pairs: one pair lived mostly in the state of Minas Gerais and the other lived mainly in the South of Brazil (states of Paraná and Santa Catarina). This study was carried out following The Code of Ethics of the World Medical Association (Declaration of Helsinki) approved by the Brazilian National Council of Ethics in Research (CONEP) and received the certification number CAAE 48609515.6.1001.5561.

2.2. Voice recordings

The twin siblings were invited to attend the Ethology lab (IPUSP) together and say, one per time (and without the presence of the co-twin in the room), the phrase "Hi, my name is Pedro" if male or "Hi, my name is Ana" if female. We isolated the diphthong [ˈOj] from the spoken sentences for this initial acoustic analysis (the phonetic analysis of the rest of the speeches will be presented in due course) due to its double vowels. Vowels have the best-defined formant structure (Kent, 2015), which could be associated with voice quality in Portuguese (Lopes et al.,

Table 1
Number of all participants according to sex and zygosity. The last line presents the number of MZ and DZ pairs.

Sex	MZ	DZ	total
F	100	24	124
M	42	6	48
Pairs	71	15	86

Note: MZ, monozygotic twins; DZ, Dizygotic twins; F, Female; M, Male participants.

2018), and using a diphthong we avoid overlapping in articulatory gestures for the consonant and vowel (Whiteside & Rixon, 2003).

Data collection was carried out in a silent room (4.5 m × 3.5 m) with closed doors. The recordist used a Zoom H1 Handy Recorder Zoom paired with a studio microphone BM8000 with high-precision electronic circuitry, designed to minimize background noise and isolate the main sound source. The microphone was positioned at a distance of 1 open palm from the participant's mouth (approximately 15 cm). Recordings were performed using 24-bit/48 kHz in stereo mode, saved in not compressed format (.wav file). All participants were seated.

2.3. Acoustic analysis

For the acoustic measurements, we used the free and open-source Praat (Boersma & Weenink, 2020). First, we extracted only one audio channel to conduct diphthong onset and end. A narrowband spectrogram was obtained from the spectrum menu using a 0.01 window length and a view range of 15,000 Hz (Barbosa & Madureira, 2015) (Fig. 1). For each diphthong, we obtained jitter, shimmer, harmonic-to-noise ratio (HNR), total duration, fundamental frequency (f0), and formants (F1 to F4) values (Table 2). To obtain these values we adopted broadband spectrogram a short-term spectral analysis which script is included in the Praat speech analysis systems. It computes the Linear Prediction Analysis (LPC) coefficients and expresses formants as their average value, in Hertz (Hz).

Twin pairs were excluded from the analysis if we detected problems in the recording, such as noise, or in the case where we had data for only one of the twins. After the exclusion of 6 twin pairs due to these recording issues, we remained with 80 pairs that were used in the comparison tests.

2.4. Statistical analysis

We first estimated the average and range values of the voice parameters of MZ and DZ males and females. Due to the low representation of DZ males, we conducted the statistical analysis only on the female sample data, providing just a description of the male parameters. We tested if there were differences between MZ and Dz sisters pairs using Welch's t-tests, which do not assume equal variances between groups. To verify the hypothesis that MZ is more alike than DZ, we tested the similarity between sisters using an Intraclass Correlation Coefficient (ICC) where values higher than 0.75 were considered a strong correlation; below 0.5, a weak correlation (Koo & Li, 2016). ICC gives how similar are the units of the same group, the closer to 1 the more similar.

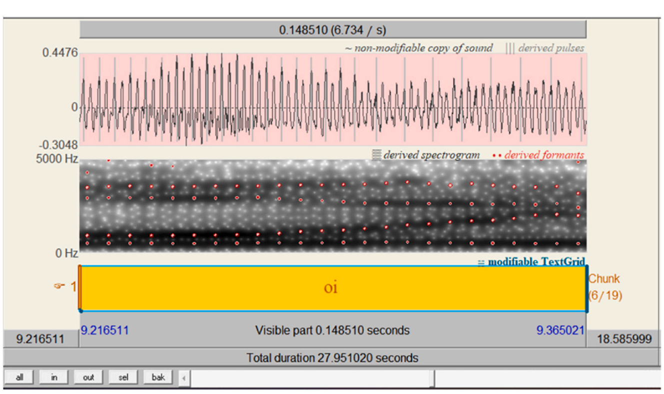

Fig. 1. A female participant [Ój] representation in Praat. The soundwave generated by the diphthong production (in the gray bar, starting at 9.216511 s and lasting 0.148 s) is presented in the Waveform (top) and the digital Spectrogram representation (in the middle). The red points in the spectrogram represent the formant contours from bottom to top (F1, F2, F3, F4). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Acoustic parameters definition and perception (Barbosa & Madureira, 2015).

Feature	Description
Duration	The length of time the diphthong lasts, here expressed in seconds.
Fundamental frequency (f0)	f0 is the resonant frequency of the sound-producing tube. It is a crucial auditory cue for intonation, tone, and lexical stress. In Praat, f0 is synonymous with pitch.
Formants (F1 to F4)	Resonance peaks (or emphasized frequencies) are generated in the filter part of the system. In the spectrogram are darker bands along the sound bandwidth.
Jitter	Measures the level of f0 perturbation, indicating the micro-instability of vocal fold vibrations. We adopt the % of jitter given by Praat
Shimmer	Measures the level of general amplitude perturbation, indicating the variability of the period amplitude of vocal fold vibration.
Harmonic to Noise Ratio (HNR)	Measures the degree of acoustic periodicity or harmonicity. Is expressed in dB: if 99% of the energy of the signal is in the periodic part, and 1% is noise, the HNR is 10 × log10 (99/1) = 20 dB. If HNR = 0 dB, harmonics and the noise have no difference in energy.

The effects of zygosity on sisters' similarity were estimated using the Euclidean distances between the twin sisters and to verify a potential relationship between zygosity, age, and the formants, we generated Generalized Linear Models (GLMM). The lowest Akaike Information Criterion (AIC) was used as the model selection criterion when comparing the models to determine which best explained the relationship between sisters' voices. We started with a complete model with age, zygosity, and its interaction, and a Tweedie distribution with a logarithmic binding function. Variations of this model were tested until an optimal model was reached, always maintaining in all models the interaction between age and zygosity.

3. Results

3.1. Description of male and female diphthong

Our sample of 160 individuals (80 pairs) produced diphthongs [Ój] lasting on average 0.24 ± 0.07s and varying from 0.09 to 0.44s according to sex and zygosity (Table 3). Fig. 2 represents the spectral parameters distribution according to sex. Note how the frequencies present a similar growing pattern in both sexes. The mean fundamental frequency of female diphthongs was 213.31 ± 25.69 Hz and the formants were 505.31 ± 43.69 Hz (F1), 1505.04 ± 166.27 Hz (F2), 2702.93 ± 196.08 Hz (F3), and 3649.97 ± 272.47 Hz (F4). Male fundamental frequency (f0) and formants were lower than females, as follows: 144.76 ± 41.03 (f0), 469.63 ± 75.25 (F1), 1424.93 ± 185.26 (F2), 2521.26 ± 200.57 (F3), and 3400.84 ± 184.31 Hz (F4).

3.2. Between sisters' similarities

From now on, the results refer only to the female sample of participants (twin sisters, MZ = 46; DZ = 10). We found no differences between DZ sisters DZ in any acoustic parameters' variance measured in Welch's *t*-test. Significant differences were found in the first formant (F1) for MZ twins ($p = 0.02$), while other parameters showed no significant differences. Notably, F1 variability was higher in MZ twins compared to DZ twins (Table 4).

Nevertheless, analyzing intra-sister similarity (Table 5) we found weak (F1, jitter, and HNR) to moderate (F3: ICC = 0.666 and F4: ICC = 0.638) correlations between MZ sisters. While DZ sisters were more likely to each other only in F1 and it was higher (ICC = 0.619) than between MZ sisters (ICC = 0.34).

Table 3
Individuals' diphthongs characterization according to mean and standard deviation (sd) of the acoustic parameters (n = 160 individuals; 80 pairs).

	Female				Male			
	MZ		DZ		Mz		DZ	
	Mean	± sd	Mean	± sd	Mean	± sd	Mean	± sd
Age	30	10.9	31	12.5	29	11.8	30	13.2
Duration (s)	0.25	0.07	0.24	0.07	0.21	0.06	0.22	0.06
f0 (Hz)	504.16	43.76	510.22	44.06	469.55	77.09	470.11	68.68
F1 (Hz)	215.01	25.93	206.04	23.80	146.37	43.90	134.61	8.49
F2 (Hz)	1508.50	166.14	1490.24	169.92	1424.97	174.29	1424.66	265.26
F3 (Hz)	2701.93	128.23	2707.18	371.08	2514.63	203.11	2563.28	195.53
F4 (Hz)	3645.88	273.12	3667.41	275.34	3400.58	191.57	3402.52	143.39
Jitter (%)	1.58	0.82	1.56	0.70	1.95	1.13	1.88	0.49
Shimmer (%)	9.19	3.06	10.27	4.57	9.43	3.16	9.84	3.57
HNR	12.18	3.11	12.20	3.15	9.75	2.31	9.04	3.71

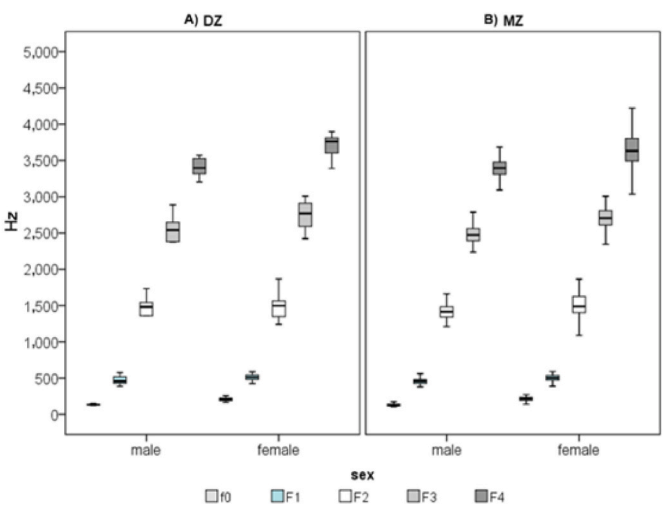

Fig. 2. Spectral parameters are presented in boxplots for DZ (A) and MZ (B) twins (n = 160 individuals; 80 pairs).

Table 4
Mean and standard deviation of sister pairs and statistical comparison in Welch's *t*-test Results for 46 pairs of MZ and DZ 10 pairs of Twins' Acoustic Parameters. Significant values are in bold.

	MZ (n = 46)			DZ (n = 10)		
	Mean	SD	T (p-value)	Mean	SD	T (p-value)
Duration	0.07	0.06	−0.21 (0.83)	0.08	0.07	1.01 (0.33)
f0	18.97	14.42	−0.95 (0.34)	22.04	17.45	1.46 (0.16)
F1	39.85	36.01	**−2.28 (0.02)**	33.86	18.05	−0.22 (0.83)
F2	146.535	108.08	−0.30 (0.76)	186.15	203.76	1.38 (0.19)
F3	138.455	118.09	1.41 (0.16)	185.85	128.01	−0.12 (0.90)
F4	200.07	171.07	0.49 (0.63)	248.62	162.79	0.26 (0.80)
Jitter	0.77	0.61	−0.58 (0.56)	0.73	0.4	0.13 (0.90)
Shimmer	3.55	2.26	−1.50 (0.14)	2.47	2.3	0.13 (0.90)
HNR	3.02	2.18	0.90 (0.37)	2.55	2.35	1.01 (0.33)

Table 5
Female Intraclass Correlation Coefficient (ICC) in acoustic parameters (MZ = 46; DZ = 10). Significant values are in bold.

Formant	zygosity	ICC	IC 95%		F	*p*
			min	max		
F1	**MZ**	**0.34**	**0.068**	**0.565**	**2.03**	**0.008**
	DZ	**0.619**	**−0.022**	**0.899**	**4.244**	**0.028**
F2	**MZ**	**0.369**	**0.101**	**0.587**	**2.17**	**0.004**
	DZ	−0.273	−0.772	0.434	0.571	0.777
F3	**MZ**	**0.666**	**0.476**	**0.797**	**4.989**	**<0.0001**
	DZ	0.054	−0.598	0.663	1.114	0.441
F4	**MZ**	**0.638**	**0.438**	**0.779**	**4.531**	**<0.0001**
	DZ	0.376	−0.336	0.814	2.205	0.142
Jitter	**MZ**	**0.374**	**0.106**	**0.591**	**2.194**	**0.004**
	DZ	−0.564	−0.882	0.105	0.279	0.955
Shimmer	MZ	0.112	−0.172	0.379	1.252	0.22
	DZ	0.306	−0.404	0.786	1.882	0.195
HNR	**MZ**	**0.366**	**0.097**	**0.585**	**2.154**	**0.004**
	DZ	0.377	−0.334	0.815	2.212	0.141

Note: Following Koo and Li (2016), ICC >0.75 represents a strong correlation, and ICC <0.5, it is weak.

3.3. *Influence of zygosity and age on spectral structure*

Since verifying the interaction between age and zygosity was the goal of our analyses, we kept this relation in all models. Age and zygosity alone did not affect the voice in GLMM. When used in interaction (zygosity × age), GLMM pointed effects on formant distances between sisters (Table 6). Estimates indicate that DZ tends to differentiate more intensely with the aging process than MZ. The higher the age, the greater the difference between DZ sisters on F1 and F3. In F1, the distance between sisters had a zygosity × age effect, Wald χ 2 (2) = 6.04 and p = 0.050 on their differentiation, being more pronounced on DZ (Fig. 3A). In F3, the model indicated an interaction effect between zygosity × age and a main effect of age, Wald χ2 = 4.074 and p = 0.044. Again, DZ seems to differentiate more as age increases than MZ (Fig. 3C).

There was no interaction effect for zygosity × age in formants F2 and

Table 6
The smaller AIC models better reflected sisters' differences based on Euclidean distance (MZ = 46, DZ = 10).

Formant	Model selected	AIC	Wald χ2	P-value
F1	Zygosity[a]age	489.05	6.04	**0.05**
F2	Zygosity[a]age [b] age	604.85	2.280	0.32
F3	Zygosity[a]age [b] age	638.34	4.074	**0.044**
F4	Zygosity[a]age	674.38	0.692	0.708

Note.
[a] Represents an interaction between factors.
[b] Represents factor combinations in tested models. Wald X^2 resulted from the test; significant values are in bold. Akaike information criterion (AIC).

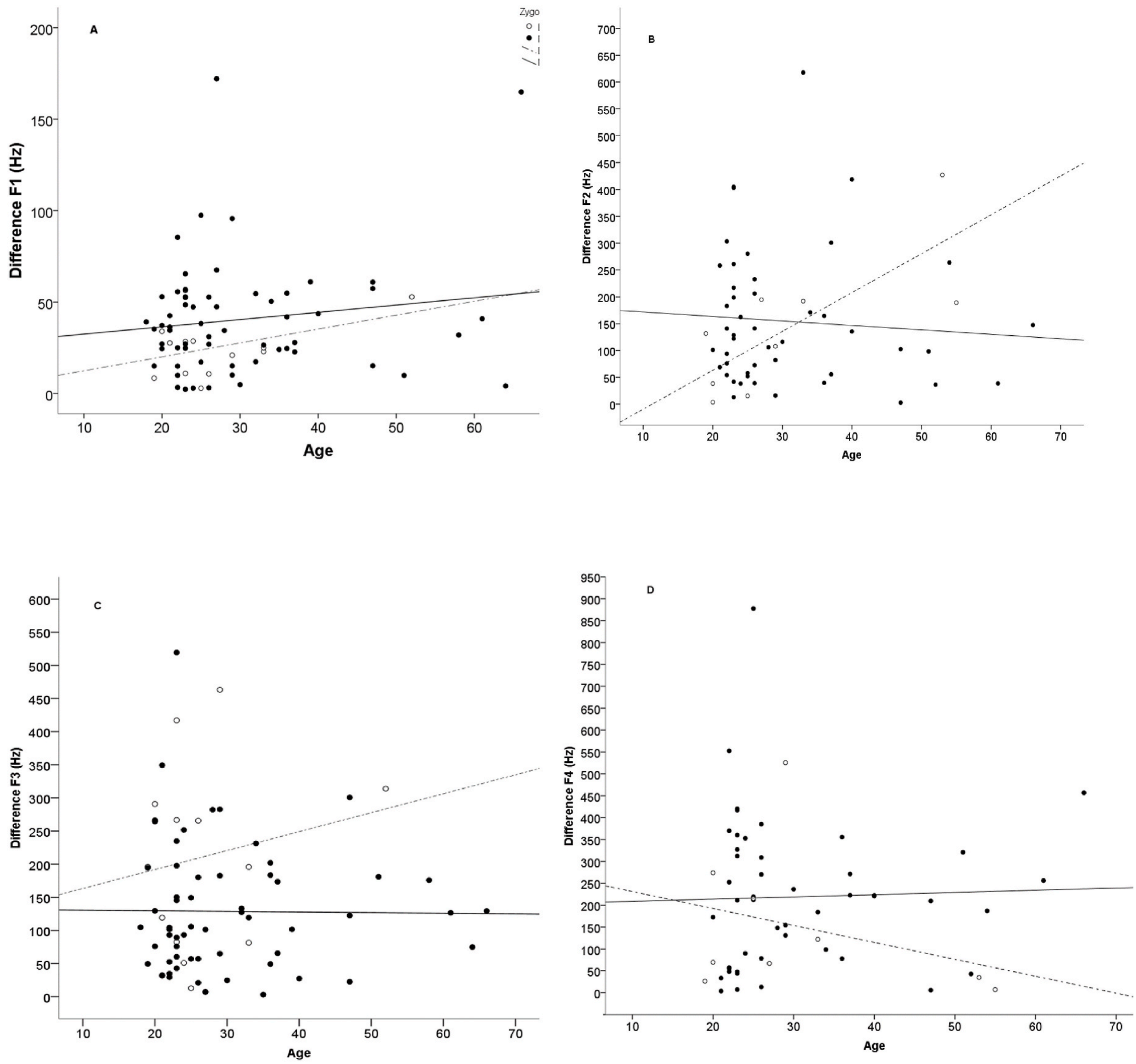

Fig. 3. Effects of the interaction age × zygosity on formants F1 (A), F2 (B), F3 (C), and F4 (D) in twin sisters' voice.

F4. We could not verify the effect on F2 formant as shown in the final model containing the interaction and the age effect, Wald χ 2 (2) = 2.280 and p = 0.320. Although, the graphical inspection indicates that this effect may exist (Fig. 3b). In the same way, we could not identify any effect on the F4, pointed out by the final model containing only the AIC interaction, Wald $\chi 2 = 0.692$, and p = 0.708 (Fig. 3 B and D).

4. Discussion

In this study, we aimed to contribute to the distinction of MZ twins by the voice which would be a cheaper and noninvasive way of determining zygosity (Ariyaeeinia et al., 2008). Intending to individualize MZ twins by voice, we started here by analyzing vowels diphthong due to their best-defined formant structure (Kent, 2015), and in adult participants. In Brazilian Portuguese, the vowels differ from consonants for being voiced, i.e., the vocal folds vibrate during the emission (Barbosa & Albano, 2004). There are seven oral vowels in the stressed position (/i e a O u) that play a role in sentence-level intelligibility (Ishikawa & Webster, 2023). When combined, vowels produce diphthongs, like the chosen ['Oj] composed by both a nasalized [O] and a not nasalized [i] form.

In short, our data showed differences only in F1 between MZ sisters, and no differences between DZ sisters, which frustrates in part our aim to contribute to individual differentiation through a short voice expression that was diphthong. It still needs to be investigated how it would work for short consonants or mixed syllables. On the other hand, the weak to moderate correlation found between MZ sisters in the four formants, jitter, and HNR should motivate further investigations around these parameters or in a larger sample of MZ twins. We choose to start with a short verbal emission, for its facility or recording and analyzing, but it may be the case that even these parameters would differ between MZ siblings in a sentence. For instance, Nolan and Oh (1996) found differences between MZ siblings in the emissions of the consonants approximant/r/and the lateral approximant/l/in reading speech.

We found no distinction in f0, not even between DZ siblings, which is comprehensive. The strong relationship between biological traits and f0 is well documented in the literature (Debruyne et al., 2002; Barbosa & Madureira, 2015). For instance, the vocal vibration rate is determined by the size and tension level of the vocal folds that have a genetic influence (Barbosa & Madureira, 2015; Debruyne et al., 2002; Przybyla, Horii & Crawford, 1992). Some differences between MZ siblings were found by Fuchs et al. (2000), Van Gysel et al. (2001), and Whiteside and Rixon (2013) but in case studies, reading or in dynamic parameters (f0 onset, offset, and semitones change).

On the other hand, f0 measures are related to sex (e.g., Debruyne et al., 2002) and it is clearly shown in Table 2. Men's reluctance to participate in research is a recurring issue (Weinrich, 1969; Butera, 2006; Thorogood et al., 1993) that reduced to 3 the DZ pairs and to 19 the MZ in our study. It may be possible to speculate based on the average values presented in Table 2, that females' diphthongs are shortly longer, have a lower jitter percentage, and have a higher HNR than the males' emission. MZ and DZ brothers were on average similar in spectral and voice perturbation parameters, except for F1 and jitter %, which were higher in MZ individuals. It may be a numeric reflection of the higher variation of MZ values around the mean or a sample bias. There was no correlation between MZ sisters in shimmer (a perturbation voice parameter), matching partially the findings of Van Lierde et al. (2005) to which both jitter and shimmer were more susceptible to environmental influences in comparison to spectral ones.

Monozygotic sisters were more alike to each other in all four formants analyzed, even though these correlations were weak or moderate. On the other hand, the DZ sisters were more alike to each other only for the F1. This could be related to anatomical aspects related to the articulatory and resonant mechanisms of the formants where the three first formants can provide information on vowel identity (Ericsdotter Nordgren, 2019). The first formant is associated with the articulatory dimension of tongue height; F2 and F3 are related to tongue backness and lip roundness respectively (Chiba & Kajiyama, 1941; Peterson & Barney, 1952; Potter & Steinberg, 1950). F1 shows higher values for open than for closed vowels, F2 and F3 to a lesser degree, show lower values for back vowels and are influenced by other resonatory settings such as lip rounding (Stevens, 1998).

Overall, our results support previous findings of the higher similarity between MZ and DZ twins, as expected (San Segundo et al., 2017). However, the current study included a more representative sample of twin pairs, including both males and females.

4.1. Aging in voice

At the beginning of their lives, identical twins share the genetic heritage that guided the development of both the vocal tract anatomy and functioning and the psychological processes acting in voice (Debruyne et al., 2002; Van Lierde et al., 2005; Weirich & Lancia, 2011). By 4 years of age, the voice incorporates the effects of sexual differentiation that usually, but not in all humans, would cause maturational reductions in formant bandwidth (Kent & Vorperian, 2018). Even the most robust parameter of the voice, the fundamental frequency (f0) seems to accumulate some differentiation in adult life as we found here and already previewed in the literature (Fuchs et al., 2000; Van Gysel et al., 2001; Whiteside & Rixon, 2013). The single interactions each sibling made with their unshared environment may develop particular habits that affect the voice, like smoking or singing practice (Behlau, 2001; Stemple et al., 2018).

Increasing age enhanced the difference between sisters, as suggested by the GLMM, but age alone did not affect the differentiation between sisters. The interaction zygosity × age affected the first and third formants, but in different directions; the higher the DZ sister's age, the greater their difference in the first formant position in the diphthong spectral structure; older MZ sisters differed in the 3rd formant.

We propose that formants may be a good approach to characterize even a short voice expression and agree with San Segundo and Yang (2019) about the unviability and nonsense of determining the extent one identifies genetic contribution that is immune from the experience promoted by the shared environment. Future analyses should better understand the Evo-Devo system in voice production in mammals. Besides genetic influence, MZ twins share cognitive and neuromuscular factors that contribute to the greater overlap in the articulation skills of MZ than DZ twins (Matheny & Bruggemann, 1973).

4.2. Limitations to the study

Choosing single-word or monosyllabic samples for test words is not new, and a neutral context for vowel articulation is recommended to reduce the effects of changes in vowel characteristics. (Chesworth et al., 2003; Robb & Chen, 2009). Once no single type of sample is ideal for all purposes (Kent & Vorperian, 2018) we chose a very short and stereotyped segment that is easy to articulate at any age. Our sample consists of 80 pairs of twins, which is substantial. However, we only analyzed one diphthong per participant, potentially introducing uncertainty in result interpretation. Our intention was not to describe human voice variability or to use it for clinical purposes, but rather to estimate intrasibling concordance.

Albeit our best efforts, it is well documented in the literature that men do not generally volunteer to participate in research studies. The limitations of this study also include the fact that the statistical analyses were restricted to women, which precludes the ability to generalize the findings to a broader population. Future studies could join efforts to recruit men to volunteer.

Despite the utilization of one brief word, we were able to identify differences and similarities between twins and the verification of the aging effect on voice parameters. Future research could be enhanced by increasing the sample size and including more than one speech token per speaker. This would facilitate the evaluation of the heritability and usability of these findings in contexts different from our own, such as clinical and technological development research.

5. Conclusion

Despite the shared gene-environmental heritage of monozygotic twins, siblings' voices may differ in shimmer and formants dispersion, suggesting the flexibility of this human behavior and the range of variance of the combinatorial possibilities produced by the vocal tract elements during verbal communication. On the other hand, the lack of MZ intrapair differences suggests that formants, jitter, and HNR, besides f0, may be useful, in the future, for zygosity estimation.

This study contributes to the understanding of voice similarities among twins and provides a foundation for future investigations into the heritability of voice parameters and its potential clinical and technological applications. Despite its inherent limitations, the methodology employed provides a robust foundation for subsequent studies seeking to expand and diversify samples and explore voice variations across different demographic contexts.

Funding

This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP (E.O. grant number 2014/50282-5 and 2020/14250-2), (L.C.L grant numbers 2020/14,250–2, 2022/08063–0) and by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES (funding code 001).

CRediT authorship contribution statement

Paula Bruna Campos: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis. **Luchesi Lilian Cristina:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis. **Lucci Tania Kiehl:** Writing – review & editing, Validation, Resources, Methodology, Data curation. **Otta Emma:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Monticelli Patrícia Ferreira:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the twins who participated in the study and the staff of the University of Sao Paulo Twin Panel (Painel USP de Gêmeos). Also, we are grateful to the corrections made by the anonymous referees and editor.

References

Alkhatib, B., Eddin, M. M. K., & Syria, D. (2023). ASR features extraction using MFCC and LPC: A Comparative study. *Journal of Digital Information Management, 21*(2), 39–49.

Ariyaeeinia, A., Morrison, C., Malegaonkar, A., & Black, S. (2008). A test of the effectiveness of speaker verification for differentiating between identical twins. *Science & Justice, 48*(4), 182–186. https://doi.org/10.1016/j.scijus.2008.02.002

Barbosa, P., & Albano, E. (2004). Brazilian Portuguese. *Journal of the International Phonetic Association, 34*(2), 227–232. https://doi.org/10.1017/S0025100304001756

Barbosa, P. A., & Madureira, S. (2015). *Manual de fonética acústica experimental: aplicações a dados do português. Cortez editora*.

Behlau, M. S. (2001). *Voz*. Revinter: O Livro do Especialista Rio de Janeiro.

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* v 6.1. 21. .

Brockmann, M., Storck, C., Carding, P. N., & Drinnan, M. J. (2008). Voice loudness and gender effects on jitter and shimmer in healthy adults. *Journal of Speech, Language, and Hearing Research, 51*, 1152–1160. https://doi.org/10.1044/1092-4388(2008/06-0208

Butera, K. J. (2006). Manhunt: The challenge of enticing men to participate in a study on friendship. *Qualitative Inquiry, 12*(6), 1262–1282. https://doi.org/10.1177/1077800406288634

Casanova, E. (2021). Deep learning approaches for speech synthesis and speaker verification. In E. Otta, & P. F. Monticelli (Eds.), *Acoustic communication: An interdisciplinary approach*. Instituto de Psicologia da Universidade de São Paulo.

Casanova, E., Shulby, C. D., & Aluísio, S. M. (2021). Deep learning approaches for speech synthesis and speaker verification. In E. Otta, & P. F. Monticelli (Eds.), *Acoustic communication: An interdisciplinary approach*. Instituto de Psicologia da Universidade de São Paulo.

Cavalcanti, J. C., Eriksson, A., & Barbosa, P. A. (2021). Multiparametric analysis of speaking fundamental frequency in genetically related speakers using different speech materials: Some forensic implications. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2021.08.013

Chiba, T., & Kajiyama, M. (1941). *The vowel: Its nature and structure*. Tokyo: Kaiseikan.

Debruyne, F., Decoster, W., Van Gijsel, A., & Vercammen, J. (2002). Speaking fundamental frequency in monozygotic and dizygotic twins. *Journal of Voice, 16*(4), 466–471.

Eichhorn, J., Kent, R., Austin, D., & Vorperian, H. K. (2017). Effects of aging on vocal fundamental frequency and vowel formants in men and women. *Journal of Voice, 32* (5), P644_E1, 644.E9 https://doir.org/10.1016/j.jvoice.2017.08.003.

Ericsdotter Nordgren, C. (2019). Phonetics of vowels. In *Oxford research encyclopedia of linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.405.

Felippe, A. C. N. D., Grillo, M. H. M. M., & Grechi, T. H. (2006). Normatização de medidas acústicas para vozes normais. *Revista Brasileira de Otorrinolaringologia, 72*, 659–664.

Fernandes, L. D. C., Polido, A., & Wertzner, H. F. (1999). Contribuições da análise acústica para o processo diagnóstico da alteração da articulação. *Pró-fono* (pp. 61–67).

Fuchs, M., Oeken, J., Hotopp, T., Täschner, R., Hentschel, B., & Behrendt, W. (2000). Similarity of monozygotic twins regarding vocal performance and acoustic markers and possible clinical significance. *HNO, 48*(6), 462–469.

Hagenbeek, F. A., van Dongen, J., Pool, R., & Boomsma, D. I. (2022). Twins and omics: The role of twin studies in multi-omics. In A. D. Tarnoki, D. L. Tarnoki, J. Harris, & N. L. Segal (Eds.), *Twin research for everyone: From biology to health, epigenetics, and psychology* (pp. 547–584). Elsevier.

Ishikawa, K., & Webster, J. (2023). The formant bandwidth as a measure of vowel intelligibility in dysphonic speech. *Journal of Voice, 37*(2), 173–177. https://doi.org/10.1016/j.jvoice.2020.10.012

Kent, R. D. (2015). Análise acústica da fala. *Trad. Alexsandro Rodrigues Meireles. São Paulo: Cortez Editora*.

Kent, R. D., & Vorperian, H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders, 74*, 74–97. https://doi.org/10.1016/j.jcomdis.2018.05.004

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15* (2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012. Epub 2016 Mar 31. Erratum in: J Chiropr Med. 2017 Dec;16(4):346. PMID: 27330520; PMCID: PMC4913118.

Lopes, L. W., Alves, J. do N., Evangelista, D. da S., França, F. P., Vieira, V. J. D., Lima-Silva, M. F. B. de, & Pernambuco, L. de A. (2018). Acurácia das medidas acústicas tradicionais e formânticas na avaliação da qualidade vocal. *Coda, 30*(5), Article e20170282. https://doi.org/10.1590/2317-1782/20182017282

Matthews, L. J., & Turkheimer, E. (2022). Three legs of the missing heritability problem. *Stud Studies in History and Philosophy of Science, 93*, 183–191. https://doi.org/10.1016/j.shpsa.2022.04.004

Moondra, A., & Chahal, P. (2023). Improved speaker recognition for degraded human voice using modified-MFCC and LPC with CNN. *International Journal of Advanced Computer Science and Applications, 14*(4), 143–151. https://doi.org/10.14569/IJACSA.2023.0140416

Nolan, F., & Oh, T. (1996). Identical twins, different voices. *International Journal of Speech Language and the Law, 3*(1), 39–49.

Otta, E., De Souza Fernandes, E., Bueno, J. A., Dos Santos, K. L., Segal, N. L., Lucci, T. K., Ferreira, I. F., Cesar, G. C., David, V. F., Tatit, D. P., Short, P. C. A., Fernandes, L. O., Crispim, A. C., Moretto, M. L. T., Andrade, N. C., Corte, S., Tobo, P. R., Barrichello, C. R., De Sousa, R. C. G., … Ribeiro, F. J. L. (2019). The university of São Paulo twin Panel: Current status and prospects for Brazilian twin studies in behavioral research. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies, 22*(6), 467–474. https://doi.org/10.1017/THG.2019.34

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*(2), 175–184.

Possani, C. (2021). Zygosity diagnosis in adult twins based on voice resemblance. In E. Otta, & P. F. Monticelli (Eds.), *Acoustic communication: An interdisciplinary approach*. Instituto de Psicologia da Universidade de São Paulo.

Potter, R. K., & Steinberg, J. C. (1950). Toward the specification of speech. *Journal of the Acoustical Society of America, 22*(6), 807–820.

Sabatier, S. B., Trester, M. R., & Dawson, J. M. (2019). Measurement of the impact of identical twin voices on automatic speaker recognition. *Measurement, 134*, 385–389. https://doi.org/10.1016/j.measurement.2018.10.057

San Segundo, E., Tsanas, A., & Gómez-Vilda, P. (2017). Euclidean distances as measures of speaker similarity including identical twin pairs: A forensic investigation using source and filter voice characteristics. *Forensic Science International, 270*, 25–38. https://doi.org/10.1016/j.forsciint.2016.11.020

San Segundo, E., & Yang, J. (2019). Formant dynamics of Spanish vocalic sequences in related speakers: A forensic-voice-comparison investigation. *Journal of Phonetics, 75*, 1–26.

Smith, C. (1970). Heritability of liability and concordance in monozygous twins. *Annals of Human Genetics, 34*, 85–91.

Stemple, J. C., Roy, N., & Klaben, B. K. (2018). *Clinical voice pathology: Theory and management*. Plural Publishing.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Taylor, A. M., & Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology, 280*(3), 221–236.

Thorogood, M., Coulter, A., Jones, L., et al. (1993). Factors affecting response to an invitation to attend for a health check. *Journal of Epidemiology & Community Health, 47*, 224–228. https://doi.org/10.1136/jech.47.3.224

Tokuda, I. (2021). The source–filter theory of speech. In *Oxford research encyclopedia of linguistics*.

Van Gysel, W. D., Vercammen, J., & Debruyne, F. (2001). Voice similarity in identical twins. *Acta Oto-Rhino-Laryngologica Belgica, 55*(1), 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*(4), 511–518.

Weinrich, S. P. (1969). Self-reported reasons men decide not to participate in free prostate cancer screening. Number 1/2003. *30*(1), E12–E16. https://doi.org/10.1188/03.ONF.E12-E16.

Weirich, M., & Lancia, L. (2011). Perceived auditory similarity and its acoustic correlates in twins and unrelated speakers. In *ICPhS* (pp. 2118–2121).

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*(4), 273–297.

Whiteside, S., & Rixon, E. (2013). Speech tempo and fundamental frequency patterns: A case study of male monozygotic twins and an age-and sex-matched sibling. *Logopedics Phoniatrics Vocology, 38*(4), 173–181. Whiteside, S. P., & Rixon, E. (2013).

Zhang, Z. (2016). Mechanics of human voice production and control. *Journal of the Acoustical Society of America, 140*(4), 2614–2635. https://doi.org/10.1121/1.4964509