PERCEIVED AUDITORY SIMILARITY AND ITS ACOUSTIC CORRELATES IN TWINS AND UNRELATED SPEAKERS

Melanie Weirich[a] *& Leonardo Lancia*[b]

[a]ZAS (Center for General Linguistics), Berlin, Germany;
[b]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
weirich@zas.gwz-berlin.de; leonardo_lancia@eva.mpg.de

ABSTRACT

We conducted an AX-discrimination task to test for differences in perceived auditory similarity in monozygotic (MZ) and dizygotic (DZ) twin pairs and unrelated speakers. In addition, we performed acoustic analyses to find acoustic correlates that explain the differences in perceived similarity.

Results indicate that unrelated speakers are significantly easier to distinguish than twins, but zygosity has no effect on perceived similarity. Moreover, pair-specific auditory similarity appears in twins and unrelated speakers and can be explained by the acoustic parameters F0, shimmer, jitter and HNR.

Keywords: perceived auditory similarity, perception test, acoustic analysis, twins

1. INTRODUCTION

Perceived auditory similarity is a crucial topic in automatic speaker recognition but also in forensic speaker identification, where the testimonies of ear witnesses or descriptions of voices are important issues [9]. The ability of discriminating voices is dependent on several factors, such as quality and length of stimuli, degree of auditory similarity between the compared voices and whether the listeners are familiar with the speakers [8]. The similarity of voices has been addressed in forensic studies by comparing unrelated speakers but also related speakers, such as twins [7].

In general it can be assumed that along with the impact of linguistic background and dialectal influence, related speakers that share physiological characteristics due to genetic similarity also have more similar sounding voices than unrelated speakers. Several studies showed that monozygotic (MZ) twins have very similar voice characteristics leading to perceived similarity [2, 11].Studies that aimed at revealing the acoustic parameters which determine perceived speaker similarity found F0 to be crucial [2]. Since F0 is influenced by organic and physiological constraints and thus is more similar in MZ than in dizygotic (DZ) twins [1], a common assumption is that perceived auditory similarity is higher in MZ than in DZ twins. Interestingly, Johnson & Azara [3] found in their study that the perceptual similarity in MZ twins was not larger than in DZ twins. Nevertheless, this result is limited because only one DZ pair participated in their perception experiment. Debruyne, et al. [1] found that variation of F0 is less physiologically determined than mean F0 since MZ and DZ twins revealed the same amount of similarity within their study. In what way the differences in F0 variation (mean F0 range) might have an effect on perceived similarity has not been established yet. In the comprehensive twin study of Van Lierde, et al. [6] different voice quality characteristics were investigated and found to be very similar within MZ twins. Thus, a strong relation to organic parameters can be assumed. However, the two parameters jitter (frequency-micro-perturbations) and shimmer (amplitude-micro-perturbations) revealed no significant correlation within the twins. These parameters might be correlated with environmental factors (e.g. state of health, anxiety or tension) and could be independent of physiological constraints but there still seems to be a lack of clarity in literature. Shimmer is known to be an acoustic correlate of perceived "hoarseness" [5]. In addition, the harmonics-to-noise-ratio (HNR) that relates the harmonic level of a signal to its noise level also correlates with perceived hoarseness and breathiness. Johnson & Azara [3] mention the factor breathiness as a possible auditory cue on perceived similarity.

The above-mentioned studies lead to the following hypotheses:

- H1) Unrelated speakers are easier to distinguish than related speakers (twin pairs) and DZ twins are easier to distinguish than MZ twins.
- H2) Mean fundamental frequency (F0) is more similar in MZ than in DZ twins due to physiological constraints.
- H3) Mean F0 range varies equally in MZ and DZ twins, since it reflects learnt language behaviour.

- H4) Voice quality parameters are crucial in perceived speaker similarity and influenced by environmental factors and thus, differ more in pairs that are easy to distinguish.

Although supra-laryngeal factors also influence perceived speaker similarity, in this paper we will concentrate on the contribution of voice quality and F0. Formant patterns and spectral characteristics of consonants are discussed in great detail in [10].

2. METHOD

To investigate perceived speaker similarity an AX same-different perception test was conducted. The acoustic stimuli consisted of only one word and were permuted which resulted in pairs of two different stimuli. The paired stimuli were part of one of the following speaker groups: same speaker (different repetitions), MZ twins, DZ twins or unrelated speakers.

Furthermore, an acoustic analysis was made to look for the effect of different acoustic parameters on perceived speaker similarity.

2.1. Subjects and stimuli

Two MZ and two DZ German twin pairs (female, between 20 and 34 years old) served as speakers. The stimuli consisted of the word /ˈvaʃə/ (‘wash’, 1st person. sg.) extracted from the carrier sentence “Ich wasche Haku/Hag(/u,i,a/) im Garten” (I wash Haku/Hag(u,i,a) in the garden). 6 repetitions were selected for each of the eight speakers, resulting in a total of 48 different stimuli. The stimuli were normalized in intensity by setting the highest amplitude of each signal to 0dB (that refers to 100%), and adjusting all other amplitude values of this signal respectively.

For the AX perception test 28 native German listeners (11 male and 17 female; average age: 29.6, SD = 6.4) were asked to judge for each stimuli pair whether it belongs to the same speaker or different speakers.

2.2. Perception test

Due to time constraints of the perception test 4 listener groups were built that rated each of the possible speaker pairs but not all repetitions of each speaker. In addition, the amount of stimuli pairs within the group of unrelated speakers was reduced. In the end, each listener rated 432 different AX stimuli pairs twice, resulting in 864 AX pairs. The perception test was run in PRAAT (version 5.1.04). Subjects listened to each presented stimulus pair once over Sennheiser HD 595 headphone in a randomized order. Directly after listening to each stimulus, they were asked to click on a button “same speaker” or “different speaker”. Each stimuli pair was presented in both possible orders (AX and XA) to factor out the effect of this potentially confounding factor.

2.3. Acoustic analysis

All acoustic analyses were conducted using PRAAT. Mean and standard deviation of the fundamental frequency (F0) were calculated over the six repetitions of each speaker. We used a normalized variation coefficient (Std_norm) that is independent of mean F0 (x) [4].

$$Std\_norm = \frac{100 \cdot s}{x} \quad (1)$$

Furthermore, voice quality parameters were measured over the vowel /a/ of the target word.

- *Jitter* is an acoustic measurement of how much a given period differs from the period that immediately follows it. We calculated the five point Period Perturbation Quotient (PPQ5).
- *Shimmer* or amplitude perturbation quantifies the short-term instability of the vocal signal. We measured the Amplitude Perturbation Quotient (APQ5).
- *Harmonics-to-Noise Ratio (HNR)* gives a measurement of perceived hoarseness and/or aspiration and is expressed in dB.

3. RESULTS

3.1. Speaker discrimination scores

The 28 listeners varied in their overall correctness scores of distinguishing speakers from 67% to 95%. On average, the listeners were able to differentiate same and different speakers in 82.8% and no effect of the listeners’ gender was found.

Figure 1: Correctness scores (in %) for 4 speaker groups (correct: left bar, false: right bar)

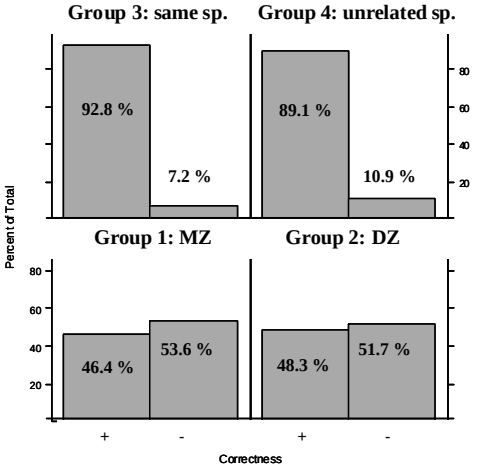


Figure 1 visualizes the correctness scores (in percent) separated by speaker groups. Group 3 (same speaker) and group 4 (unrelated speakers) reached very high correctness scores of around 90%. For the twin groups the percentage of correct identification scores reached less than 50%, supporting H1, the difficulty of distinguishing related speakers. The tendency for MZ pairs being more difficult to distinguish than DZ pairs turned out to be less than expected (46% correct answers for MZ vs. 48% correct answers for DZ pairs).

3.2. Statistical analysis

To look for a significant effect of the four different groups on the correctness scores, we calculated a generalized linear mixed model (family = *binomial)* in R (version 2.9.0). The factor GROUP was taken as fixed factor with the different speaker groups as different levels. The level same speaker was excluded from the analysis since it only served as a control group. According to H1 and the expected correctness scores, the factor GROUP was considered as an ordered factor (MZ < DZ < unrelated) and expressed through a successive difference contrast. The factors LISTENER, SPEAKER PAIR and STIMULI served as random factors.

As expected, the difference between unrelated pairs and DZ pairs turned out to be highly significant ($z = -5.117$, $p < 0.001$), but no significant difference was found between MZ and DZ pairs ($z = -0.220$, $p = .83$). Therefore, a higher perceived similarity in MZ than DZ twins could not be confirmed.

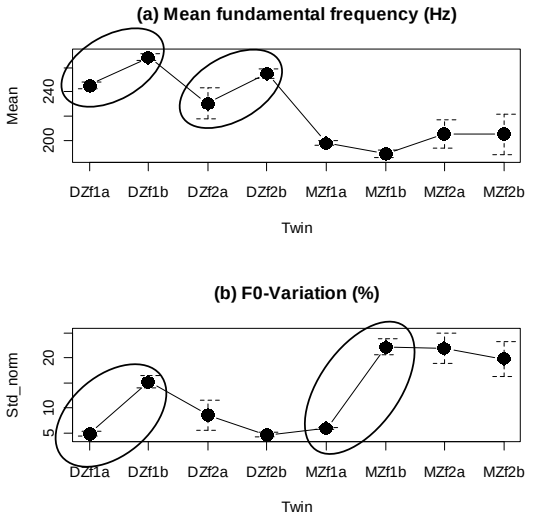
One MZ pair (MZf2) and one DZ pair (DZf1) had higher percentages of false answers (68% and 62% resp.) than the other two pairs (41% for DZf2 and 39% for MZf1). Additionally, a high amount of inter-pair variability was found between the 24 different unrelated speaker pairs. Thus the question arises, as to which acoustic correlates are responsible for the pair-specific similarity.

3.3. Acoustic correlates

3.3.1. Twins

The upper part of Figure 2 visualizes the mean fundamental frequency (F0_Mean) for each speaker. The DZ twins have higher fundamental frequencies than the MZ twins. This might be explained by their younger age (both DZ twins were 20 years old, the MZ pairs were 34 (MZf1) and 26 (MZf2) years old), but it could also be due to common inter-speaker variation. More interestingly, the difference between speakers within pairs is higher for DZ than for MZ twins, supporting the assumed influence of physiology on F0. However, the difference in F0 was not a strong factor on perceived similarity given that the MZ twins were not easier to distinguish than the DZ twins (H2).

Figure 2: Mean F0 in Hz (upper figure) and mean normalized variation in F0 in percent (lower figure) for all speakers, biggest differences marked by circles



The lower part of Figure 2 shows the mean F0 range within the sequence /ˈvaʃə/ (in %). Some speakers reveal quite high variation coefficients of more than 20%, some show less than 10%. Most inter-speaker variation was found for the pairs DZf1 and MZf1. Since DZf1 revealed very low discrimination scores a minor influence of F0-variation was attributed to perceived similarity (what might be explained by the short duration of the used stimulus). In addition, no effect of zygosity could be found, since the DZ twins do not vary more in F0 variation than the MZ twins. Thus, the findings corroborate H3 and the minor impact of biology on variation in F0 (within one word) in (normal) speakers.

Table 1 shows *shimmer* (APQ5), *jitter* (PPQ5) and *HNR* values for each speaker. Welch Two Sample t-tests were conducted. MZf1 revealed differences in all voice quality measures. Speaker MZf1a has a very low HNR value of 9.83 dB which is associated with perceived hoarseness. Her jitter and shimmer values are significantly higher than her sister's. Taken together, this leads to the impression of a hoarse and breathy voice and explains the results of the perception test: MZf1 was the twin pair that was confused the least. Hence, voice

quality parameters are crucial in distinguishing similar sounding voices like the voices of twins and can be influenced by environmental factors (MZf1a is a teacher).

Table 1: Mean values for APQ5, PPQ5 and HNR, significant differences in bold ($p < .05$ for APQ5 and PPQ5, $p < .01$ for HNR

Twin pair	Twin	APQ5	PPQ5	HNR
MZf1	MZf1a	**0.0351**	**0.0060**	**15.84**
	MZf1b	**0.0555**	**0.0125**	**9.83**
MZf2	MZf2b	0.0470	0.0061	12.00
	MZf2a	0.0365	0.0044	13.62
DZf1	DZf1a	0.0320	0.0069	**15.37**
	DZf1b	0.0269	0.0041	**18.30**
DZf2	DZf2b	0.0256	0.0041	14.44
	DZf2a	0.0359	0.0039	14.01

3.3.2. Unrelated speakers

Differences in discrimination scores were found also between the 24 unrelated speaker pairs. Pearson correlations were calculated between discrimination scores and differences in acoustic parameters. If speakers differ to a large extent in an acoustic parameter, the error score (and hence the perceived similarity) should be small, given that this parameter has an influence on the perceived speaker identity.

Table 2: Correlations between perceived similarity and acoustic parameters in unrelated pairs, significant correlations in bold ($p < 0.05$)

	Δ_F0	Δ_SD	Δ_apq5	Δ_ppq5	Δ_HNR
R	-0.61	-0.27	-0.46	-0.27	-0.39
R^2	**0.37**	0.07	**0.21**	0.07	0.16
p	**0.001**	0.19	**0.02**	0.19	0.05

In Table 2 it can be seen that correlations for all acoustic parameters are negative, which is an indication of the expected negative effect of acoustic differences on perceived similarity. However, only F0 ($t = -3.6$, $df = 22$) and APQ5 ($t = -2.4$, $df = 22$) show a significant correlation ($p < .05$). HNR marginally fails to show significance with $p = 0.054$ ($t = -2.0$, $df = 22$). The strongest impact factor on perceived similarity turns out to be F0 ($R^2 = 0.37$), revealing that over one third of the variance in error scores can be explained by the difference in F0. Interestingly, F0 was not the major factor when distinguishing the voices of twins, since the MZ twins were more similar in their F0 than the DZ twins, and the DZ twin pair with significant differences in F0 (DZf2) was mixed up more often then one of the MZ pairs (MZf1). Hence, F0 seems to be crucial when comparing unrelated speakers but it does not seem to be important when comparing similar-sounding voices like that of siblings. Here, other parameters like voice quality become significant. Differences in F0 variation might play a minor role when comparing speakers with stimuli that consist of only one word.

4. CONCLUSION

Listeners succeed in differentiating unrelated speakers even if they have very little evidence to go by (only one word) but fail to distinguish twins. No influence of zygosity on perceived auditory similarity was found. Voice quality parameters seem to be helpful for very similar-sounding voices, while F0 seems to be the most important factor to distinguish unrelated speakers.

5. ACKNOWLEDGEMENTS

This work was supported by the German Federal Ministry for Education and Research (BMBF) (Grant Nr. 01UG0711). Thanks to Ralf Winkler for scripting advice.

6. REFERENCES

[1] 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.

[2] Decoster, W., Van Gysel, A., Vercammen, J., Debruyne, F. 2001. Voice similarity in identical twins. *Acta Oto-Rhino-Laryngologica Belgica* 55, 49-55.

[3] Johnson, K., Azara, M. 2000. *The Perception of Personal Identity*. Unpublished Manuscript, *http://www.linguistics.berkeley.edu/~kjohnson/papers/twinPerc.pdf*

[4] Künzel, H.J. 1987. *Sprechererkennung: Grundzüge forensischer Sprachverarbeitung.* Kriminalistik Verlag, Heidelberg.

[5] Lieberman, P. 1963. Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *J. Acoust. Soc. Am.* 35, 344-353.

[6] 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.

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

[8] Rose, P. 2002. *Forensic Speaker Identification.* Forensic Science Series. London; New York: Taylor & Francis.

[9] Schiller, N.O., Köster, O. 1996. Evaluation of a foreign speaker in forensic phonetics: A report. *Forensic Linguistics* 4, 1-17.

[10] Weirich, M. Forthcoming. *The Influence of Nature and Nurture on Speaker Specific Parameters in Twins' Speech. Acoustics, Articulation and Perception.* Dissertation, Humboldt-Universität zu Berlin, Germany.

[11] Whiteside, S.P., Rixon, E. 2000. The Identification of twins from pure (single speaker) syllables and hybrid (fused) syllables: An acoustic and perceptual case study. *Perceptual and Motor Skills* 91, 933-947.