UCLM

Ediciones de la Universidad de Castilla-La Mancha

Estudios lingüísticos de jóvenes investigadores

Pilar Morales Herrera
Pilar Peinado Expósito
Yoana Ponsoda Alcázar (Coord.)

174
colección
estudios

ESTUDIOS LINGÜÍSTICOS DE JÓVENES INVESTIGADORES

ESTUDIOS LINGÜÍSTICOS DE JÓVENES INVESTIGADORES

Coordinadoras:
Pilar Morales Herrera
Pilar Peinado Expósito
Yoana Ponsoda Alcázar

Ediciones de la Universidad
de Castilla-La Mancha
Cuenca, 2021

Edita: Ediciones de la Universidad de Castilla-La Mancha

Colección ESTUDIOS n.º 174

Materia: Lingüística histórica y comparada.
Thema: CFF

Unión de Editoriales Universitarias Españolas

Esta editorial es miembro de la UNE, lo que garantiza la difusión y comercialización de sus publicaciones a nivel nacional e internacional.

Este libro ha sido sometido a evaluación externa y aprobado por la Comisión de Publicaciones de acuerdo con el Reglamento del Servicio de Publicaciones de la Universidad de Castilla-La Mancha.

ISSN: 2255-2618
I.S.B.N.: 978-84-9044-482-5
D.O.I.: http://doi.org/10.18239/estudios_2021.174.00

Composición: Compobell
Hecho en España (U.E.) – *Made in Spain (E.U.)*

COMITÉ CIENTÍFICO

ÍNDICE

http://doi.org/10.18239/estudios_2021.174.01

TESTING THE ROLE OF LANGUAGE FAMILIARITY IN AURAL PERCEPTUAL RECOGNITION: A PILOT STUDY

JOSÉ VICENTE BENAVENT CHÁFER
Universitat de Barcelona (UB)

1. INTRODUCTION

In the field of authorship attribution, recognition tasks can be performed by using either a piece of writing or a chunk of speech from the suspect to identify. Concerning the latter, research on forensic phonetics has proven that a remarkable margin of error may arise in automatic and semi-automatic speaker recognition systems when confronted with adverse acoustic conditions, be it with telephone transmissions and background noises (Alexander et al. 2004), voice disguising through mouth masks, whispers, and raised/lowered pitch (Zhang and Tan 2008), or the fitness of automatic systems to individualised phonological traits (González-Rodríguez 2014). Instead of advocating for the aforementioned automated methods, this research conducts a series of voice line-ups for the sake of gauging the human auditory system's potential to recognise unknown voices in familiar/unfamiliar languages. In doing so, volunteers from Spanish and British universities with various sociolinguistic profiles performed as jurors, who were exposed to Spanish, English, and Dutch input with and without background noises.

The rationale behind this study lies in the premise of preventing «flawed identification procedures […] producing unreliable evidence» (Broeders and van Amelsvoort 2001: 238), which may result in wrongful convictions due to the misleading sequencing of line-ups and judicial sentences being based only on evidence derived from eye/earwitness identification procedures. Although applied guidelines recommend that voice line-ups should be constructed around the voice descriptions provided by the victims/witnesses alongside the counseling of a trained phonetician (Broeders and Van Amelsvoort 2001: 241), a recent study confirmed that eliciting such explicit requests from the witnesses/victims is proven unreliable beyond the native/non-native distinction (Tompkinson and Watt 2018: 35). It is therefore imperative to optimise the retrieval of implicit traces in the memory through the improvement of voice line-ups.

Even though successful identification seems plausible when all voice samples share the same sociolinguistic background (Broeders et al. 2002: 111), it should be warned about the limitations of both the voice line-up as a method, and the witnesses/victims as reliable sources of information, since uncontrolled factors such as the memory or traumatic stress may hinder voice perception and recognition (Hollien 2016: 13). Although these components are indeed beyond the scope of this project, the proposed voice line-up attempts to recreate a more realistic scenario whereby semi-spontaneous speech is displayed and environmental noises are heard, too. This approach attempts to move away from the traditional controlled laboratory setting, which tends to yield better overall results than those expected in a real-life case (Manzanero and Barón 2017: 59).

2. THEORETICAL FRAMEWORKS

2.1. Variationist sociolinguistics/phonetics

Up until the late 1960s' research tradition, external (gender, age, etc.) and internal factors (word stress, word order, etc.) were treated separately in how language change and variation was conceived, although the so-called Labovian paradigm demonstrated not only that language change and variation is systematic at all levels, but also that the aforementioned factors do shape and motivate said linguistic variation (Weinreich et al. 1968: 188). Consequently, these foundations give rise to the notion of idiolect, a unique stylistic imprint inherent to the language user himself/herself, thus enabling a forensic linguistics approach on potential identification and discrimination of speakers, as said variation occurs at the syntactic, semantic, discursive, pragmatic, and phonetic level (Dittmar 1996: 111).

The focal point in this study therefore remains on between-speaker differentiation, or 'inter-speaker variability' (Rose 2002: 10), from a phonetic angle. More specifically, the fundamental frequency (henceforth referred to as F0) has been employed in forensic phonetics research as a fairly stable unit of phonetic measurement, as it alludes to the distinctive vocal cords' vibration involved in speech production (Loakes 2006: 205). When handling voice samples in experimental settings, however, it is worth to note that the F0 can be altered through technical, physiological (age, illness, etc.), and even psychological factors (excitedness, time of the day, etc.) (Braun 1995: 11-14), which could end up rendering distorted audio material. As a precautionary measure in acoustic analyses, it is advised to search for phonetic units which are easy to extract, stable and recurrent throughout the voice sample, and resistant against voice masking/disguising (Nolan 1983: 11), whilst also reminding phoneticians who act as expert witnesses in court that the ensuing pieces of evidence produced through the conducting of acoustic analyses must be presented as a probabilistic claim, rather than a definite verdict, lest it fall for the 'infallibility trap' (Rose 2002: 53).

2.2. Voice line-ups

With the purpose of aiding the judicial system, speaker recognition tests such as voice line-ups/parades have been broadly accepted and used in cases where the victim/witness could not maintain a visual contact with the suspect/offender, but perceived his/her voice (San Segundo 2014). The procedure consists of «putting together an audio tape which contains recordings of a number of speakers, including the suspect» (Butcher 1996: 97). After the witness/victim is instructed on the procedure, he/she is requested to identify the suspect, although discriminating the mock speakers is equally decisive, for it prevents miscarriages of justice from occurring.

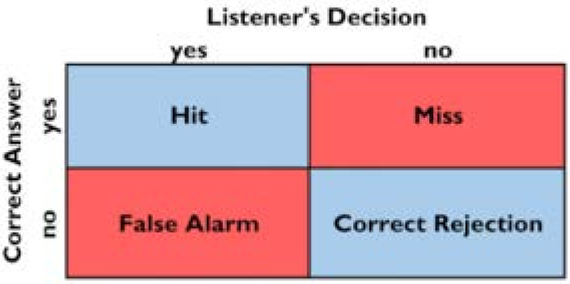

Figure 1. Possible outcomes of a speaker identification experiment (Braun 2016 :63).

The possible outcomes of a speaker identification experiment are depicted in *Figure 1*, namely 'Hit' (the juror/witness correctly identifies the suspect), 'False alarm' (the juror/witness wrongly identifies a foil speaker as the culprit), 'Miss' (the juror/witness wrongly assumes that the suspect is absent), and 'Correct rejection' (the juror/witness rightfully acknowledges that the intended suspect is absent from the line-up) (Braun 2016: 63). By heeding the expert practitioners' considerations and suggestions, this study applies the extant voice line-ups guidelines thereof (Broeders and van Amelsvoort 2001; Hollien 2012; De Jong-Lendle et al. 2015): voice line-up size of 5-7 recordings, voice samples lasting for less than 20 sec., foils and suspects should share traits such as gender, culture, and sociolinguistic background. Also, it follows checking for similar acoustic conditions in every recording of the line-up, showing no disruptive behaviours amongst the informants, and providing at least one foil whose voice is noticeably similar/dissimilar from the suspect's. On the jurors' side, they are properly instructed on the procedure and warned about the possibility of an absent suspect to replicate a real case scenario.

3. OBJECTIVES

This paper sets out to test whether a reduced familiarity with the language exposed hinders identification accuracy, as previous studies corroborated by using different combinations of jurors and stimuli (Köster et al. 1995; Köster and Schiller 1997; Hollien 2002). Secondly, it also examines whether/if the addition of background noises is correlated with a higher production of false alarms (Alexander et al. 2004), as well as figuring out to what extent recognition tasks are influenced by sociolinguistic factors (age, gender, educational level, etc.). As a final note, it seeks to further refine the already existing guidelines on voice line-ups (Broeders and van Amelsvoort 2001; Hollien 2012; De Jong-Lendle et al. 2015) to prevent the misuse of potential evidence leading to miscarriages of justice.

4. METHODOLOGY

4.1. Participants

The selected participants took part in the ad hoc on-line perception surveys sent via the department e-mail of their respective home institutions. These are the universities of Seville and Valencia (Spanish group) on the one hand, and, on the other, the universities of Bangor, Swansea, Cardiff, Roehampton, and Winchester (British group)[1]. As said jurors are enacting the role of victims/witnesses, the criterion of evoking the emotional state at the time of the incident (Rodríguez Bravo et al. 2003: 33) is fulfilled inasmuch as it refers to the state before the assault/crime (enhanced security due to the comfort of their homes), for any attempt at triggering an emotional state inherent to traumatic events would breach the code of ethics of any research activity.

1 Note that a greater number of British universities were involved due to the shortage of participants in this group. Similar to the Spanish group's distribution, the British jurors are defined by two sociolinguistic and geographical areas: Wales (Swansea, Bangor, and Cardiff), and South East England (Winchester and Roehampton).

Language test	Experimental condition	Identification scores (Points for each correct answer)	
		Spanish group (48 participants)	**British group** (31 participants)
Spanish test	1. Target-present	1	1
Familiar/Learned	2. Target-absent (with background noises)	1	1
English test	1. Target-present	1	1
Familiar/Learned	2. Target-absent (with background noises)	1	1
Dutch test	1. Target-present	1	1
Unknown	2. Target-absent (with background noises)	1	1
	Total	6	6

Table 1. Groups of jurors alongside languages tested, experimental conditions, and scores.

As *Table 1* suggests, there are three distinct types of languages being tested against two groups of jurors, namely the Spanish and British hearers. Therefore, the exposed languages (input) are coded in relation to the jurors' familiarity (familiar, learned, and unknown language). The adjacent column indicates the sub-types of language tests created according to the experimental condition: the first one recreates a scenario where the suspect to identify (the target) is among the constituents of the voice line-up, and there are no audible noise disturbances on the audio recordings. Conversely, the target is missing in the second experimental condition, and background noises are also added for an increased difficulty. Lastly, 'Identification scores' serves as a dummy variable to gauge how/if sociolinguistic variables affect the jurors' performance in said language tests (see '4.4. Statistical analysis' for further details).

4.2. Corpora

The informants' recordings were extracted from the following corpora: ESLORA (Spanish corpus), British Library Sound Archive (English corpus), IFADV (Dutch corpus). Only the Spanish corpus required a signed agreement in order to be granted access to the audio material provided, whilst the English and Dutch corpus' recordings can be freely consulted without formal requests. The available conversations range from semi-directed interviews to spontaneous exchanges. Every corpus employed in this paper complies with all the technicalities concerning the sharing of sociolinguistic features, acoustic conditions, and all the suggestions exposed at '2.2 Voice line-ups', since all recordings were conducted in the same room under the same conditions.

4.3. Perception surveys

The employed on-line perception surveys were created and distributed through Google Drive, and they adopt a between-groups experimental design. In other words, it revolves around two different groups' overall scores with the intended modification of a variable, or experimental condition (Rasinger 2013: 41). As mentioned above, said conditions refer to the presence/absence of both the target to identify and background noises. For the second condition, however, background noises (rainfall) display varying degrees of intensity across the three language tests (familiar, learned, and unknown language), as the juror may develop an 'artefact', which is the spoiling of results caused by respondents not reacting to the stimuli itself but to the task at hand, which in itself poses significant issues around the validity of the data obtained (Rasinger 2013: 43). Furthermore, the selection of foil speakers and suspects among the voice samples available was made according to the voice line-ups applied guidelines (see '2.2 Voice line-ups'). Once cleared the selection process, an acoustic criterion is set to present the suspect to identify by choosing excerpts of their speech with falling intonation, as opposed to the other voices in the line-up (foil speakers and suspects, uttering different sentences), who are introduced with instances of rising intonation patterns.

As for the explicit data on the jurors' profile obtained from the questionnaire, the corresponding sociolinguistic features are laid out in *Table 2*:

Variables	Options
Age	18-22 23-27 Over 28
Gender	Male/Female
Educational level (Studies)	Up to BA MA PhD
Musical training	Yes/No
Familiarity with linguistics/phonetics	With linguistics and phonetics With linguistics With phonetics No previous knowledge
City	Valencia/Seville Cardiff/Bangor/Swansea/Roehampton/Winchester

Table 2. Perception survey's sociolinguistic features.

4.4. Statistical analysis

The software SPSS is consulted to undertake the required statistical tests. The main objective of differentiating between familiar-learned-unknown language aural perception across the British and Spanish groups is tackled through conducting a set of chi-square tests, with the additional aid from contingency tables to visualise the existing correlations amongst the categorical variables examined. A Wilcoxon signed-rank test follows to spot significant deviation of values in each pair of language tests, that is, between the two experimental conditions mentioned above. Lastly, the variable 'identification scores' is defined as the dependent variable, whereas the sociolinguistic variables shown in *Table 2* perform as independent variables. Hence, a fixed effects model is performed to account for the variance of 'identification scores' explained by the above mentioned sociolinguistic variables, or predictors.

5. DATA ANALYSIS AND DISCUSSION

In the upcoming sections, hypotheses are formulated based on the objectives set according to previous studies and research tradition. An analysis and discussion of the results obtained shall follow thereafter.

5.1. Language familiarity

The first hypothesis is therefore formulated: 'Performance in recognition tasks is enhanced as the familiarity of the juror with the language exposed increases'. To address this matter in an orderly manner, the data obtained is introduced through line graphs. Afterwards, every case scenario is scrutinised individually: The Spanish group is examined first (1^{st} experimental condition-2^{nd} condition), and the British group shall adopt the same pattern.

5.1.1. Overview

Contrary to the literature around (foreign) speaker perception and recognition, there is no observable decreasing slope as the linguistic input becomes more unfamiliar to the juror. In fact, both the familiar and unknown language share similar success rates (hits) in both cases. There is, however, a dramatic increase in the production of false alarms in the Spanish group's L1 test, while an opposite reaction is observed in the British group's L1, whose hits rise slightly above the familiar and unknown language tests.

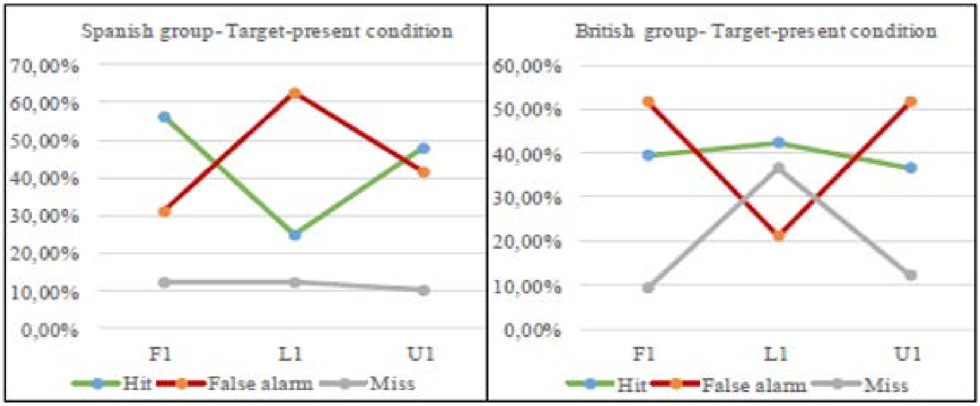

Figure 2. Voice line-up results in the first experimental condition for the Spanish and British group.

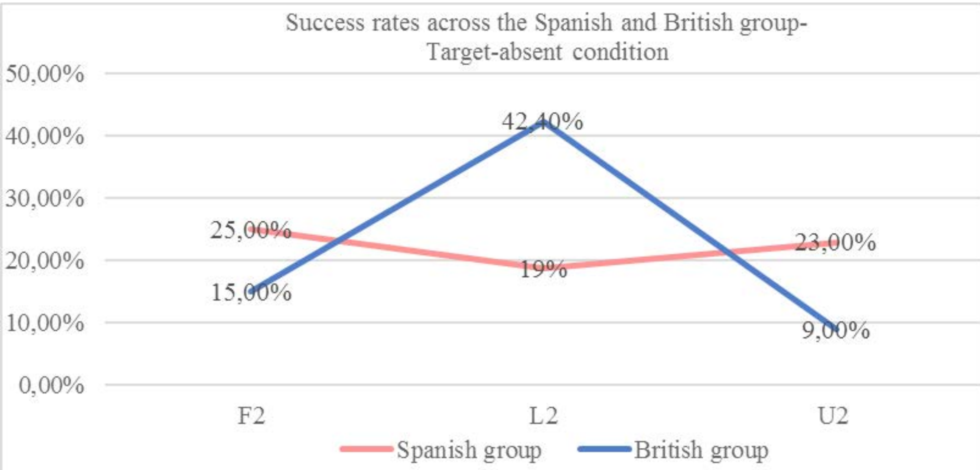

Figure 3. Voice line-up results in the second experimental condition for the Spanish and British group.

Albeit below the 50%, the scores remain stable across all language tests in the Spanish group, with a moderate decline in L2's scores. In a similar vein, the British group F2 and U2 do not differ significantly, but the L2 excels above the other tests, thus reaching its peak at 42,40% success rate.

5.1.2. Spanish group-1st condition

			Language1		
Response1	Hit	Count	27a	12b	23a b
		Adjusted Residual	2,3	-3,1	0,8
	False alarm	Count	15a	30b	20a b
		Adjusted Residual	-2,4	3,0	-0,6
	Miss	Count	6a	6a	5a
		Adjusted Residual	0,2	0,2	-0,4

Table 3. Spanish group- Language1*Response1 cross-tabulation. Cells displaying the same colours/letters indicate no significant difference.

A chi-square test of independence has found a significant correlation between the categorical variables ‘Language1’ and ‘Response1’ (X^2 (4, N = 144) = 11,341, p < 0,05). After applying a Bonferroni correction, familiar and learned language tests appear to differ notoriously in terms of hit and false alarm rates, whereas the unknown language test scores seem to fall halfway between the two extremes. The adjusted residual reflects a tendency in the learned language to reduce the chances of hits (-3,1), and increase the false alarm rates (3,0). An opposite trend is observed in the familiar language test, which seems prone to produce higher hit rates (2.3) and lesser false alarms (-2,4). As for the probabilities of missing the target in the voice line-up, no significant differences are attested amongst the three language tests.

5.1.3. Spanish group-2nd condition

Regarding the second experimental condition in the Spanish group, a chi-square test of independence did not reveal any significant correlation between the categorical variables ‘Language2’ and ‘Response2’ (X^2 (2, N = 144) = 0,563, p > 0,05).

5.1.4. British group-1st condition

The British group is hereby analysed, starting with the first experimental condition:

			Language1			
			Familiar	Learned	Unknown	
Response1	Hit	Count	13a	14a	12a	
		Adjusted Residual	0,0	0,4	-0,4	
	False alarm	Count	17a	7b	17a	
		Adjusted Residual	1,4	-2,9	1,4	
	Miss	Count	3a	12b	4a	b
		Adjusted Residual	-1,8	3,1	-1,3	

Table 4. British group- Language1*Response1 cross-tabulation. Cells displaying the same colours/letters indicate no significant difference.

A chi-square test of independence identified a significant correlation between the categorical variables 'Language1' and 'Response1' (X^2 (4, N = 99) = 12,716, p < 0,05). A Bonferroni correction is applied for every pairwise comparison, and the resulting output reveals no significant differences in the amount of hits across the three language tests. When dealing with false alarms, the learned language test surpasses the critical values with negative numbers (-2,9), unlike its familiar (1,4) and unknown (1,4) counterparts. Likewise, the amount of misses generated in the learned language are placed beyond the expected (3,1), even though the unknown language values fall somewhere between the familiar and learned language test in this respect.

5.1.5. British group-2nd condition

			Language2		
			Familiar	Learned	Unknown
Response2	Correct rejection	Count	5a	14b	3a
		Adjusted Residual	-1,2	3,4	-2,2
	False alarm	Count	28a	19b	30a
		Adjusted Residual	1,2	-3,4	2,2

Table 5. British group- Language2*Response2 cross-tabulation. Cells displaying the same colours/letters indicate no significant difference.

A chi-square test of independence has detected a significant correlation between the categorical variables 'Language2' and 'Response2' (X^2 (2, N = 99) = 12,039, p < 0,05). Once applied the Bonferroni correction, the learned language differs from the familiar and unknown language in the production of both correct rejections and false alarms. Much in line with the first experimental condition with the British participants, the learned language test exhibits greater overall performance, with higher correct rejections (3,4) and lower false alarms (-3,4) in relation to the familiar and unknown language tests.

5.2. Background noises and false alarms

The second objective, based on previous research, formulates the following hypothesis: 'Noisy conditions do hinder voice recognition, thus increasing the rate of false alarms'.

	Spanish group			British group		
	F1-F2	L1-L2	U1-U2	F1-F2	L1-L2	U1-U2
Z-score	-3,000	-0,775	-2,449	-2,828	0,000	-2,324
Asymp Sig. (2-tailed)	**0,003**	0,439	**0,014**	**0,005**	1,000[2]	**0,020**
a. Wilcoxon Signed Ranks Test						

Table 6. Pairwise comparisons on language tests' scores across two experimental conditions in British and Spanish jurors. Significant values are marked in bold (α = 0,05).

A Wilcoxon Signed Ranks tests yielded statistically significant results for the familiar (F1-F2) and unknown (U1-U2) language test pair in both groups. The learned language test comparison (L1-L2) does follow the same trend with a negative correlation, albeit without reaching statistical significance.

5.3. Sociolinguistic predictors

The third hypothesis is worded as follows: 'Sociolinguistic features do have an impact upon speaker recognition'. A fixed effects model found no significant predictors when computing the Spanish and British group scores altogether, thus rejecting the hypothesis. Nevertheless, results change drastically when isolating both groups:

Spanish group			British group		
Identification scores					
Variable	P-value	Relationship	Variable	P-value	Relationship
Studies	**0,025**	PhD≥ MA≥ Up to BA	Studies	**0,010**	PhD≥ MA≥ Up to BA
Gender	**0,057**	Male>Female	Gender	**0,040**	Female>Male
Music	**0,058**	MusicTraining> NoMusicTraining	Music	0,73	-
Age	0,27	-	Age	**0,027**	18-22> 23-27> Over 28
Linguistics	0,48	-	Linguistics	0,064	-
City	0,47	-	City	0,19	-

Table 7. Identification scores' variance explained by sociolinguistic variables in Spanish and British participants. Values in bold letters indicate significant or near-significant correlations (α = 0,05).

2 Due to the excessive influence that 'Miss' exerts upon the overall balance on L1 scores (see *Figure 2*), it is decided to account only for hits/correct rejections and false alarms, for a fair comparison of language tests. Hence the reason for a perfect correlation between the British L1 and L2.

At first glance, the variable ‘Studies’ is remarkably influential in both groups, since identification scores tend to improve as jurors’ academic level increases. Regarding ‘Gender’ in the Spanish group, albeit slightly below the 95% confidence interval, males’ performance seems enhanced the most in relation to their female counterparts. By contrast, the British group displays the opposite effect when it comes to ‘Gender’. Furthermore, having undergone musical training appears beneficial in the Spanish group (with a p-value close to the significance level), whilst its effect is nil in the British group. As for ‘Age’, no statistically significant correlation is observed in the Spanish group, whereas younger British jurors produce better overall results in speaker recognition, as opposed to the older generations. Being familiarised with linguistics and/or phonetics has no discernible effect on jurors’ performance in either group. Similarly, results remain unaffected by whichever city is chosen within each group of jurors.

6. CONCLUSIONS

First and foremost, findings seem to suggest that a reduced familiarity with the exposed language does not necessarily entail poorer results. Indeed, the best results originated from the Spanish group F1 (Hit- 56,25%/ False alarm- 31,25%) and U1 (Hit- 47,9%/ False alarm- 41,7%). Nevertheless, learned languages seem to yield disparate results in this respect. This could be due to differences in the mental processing, coding, and storage of acquired/non-acquired phonemes in contrast with those being learned. As for the acoustic conditions, background noises do impact negatively upon speaker identification tasks, regardless of the group surveyed. Learned languages appear to be less susceptible to such disturbances. As far as sociolinguistics features are concerned, only the predictor ‘Studies’ is consistent across the two groups of jurors. Thus, aural perceptual recognition seems enhanced the most as the educational level increases.

As a final remark, the results of this pilot study seem to warrant further research on improving the application of voice line-ups, as the best case scenario (target-present) yielded slightly unbalanced overall figures (≈ 40% Hits/45% False alarms), whereas perception appeared greatly hindered in target-absent tests (≈ 20% Hits/80% False alarms). In the light of the preliminary findings and correlations obtained, and however influential the small-scale size of this project might be, an extension thereof is deemed of relevance. Further research shall address and validate, whenever possible, the above mentioned claims on aural perceptual speaker recognition.

BIBLIOGRAPHIC REFERENCES

ALEXANDER, Anil, Filippo BOTTI, Damien DESSIMOZ, and Andrzej DRYGAJLO (2004): «The effect of mismatched recording conditions on human and automatic speaker recognition in forensic applications». *Forensic Science International*, 146(1), 95-99.

BRAUN, Almut. (1995): «Fundamental frequency- How speaker-specific is it?», in Almut Braun and Köster, Jens-Peter (eds.): *Studies in Forensic Phonetics*. Trier: Wissenschaftlicher Verlag, 9-23.

BRAUN, Almut. (2016): *The speaker identification ability of blind and sighted listeners: an empirical investigation*. Wiesbaden: SRINGER, 63-66.

BROEDERS, Ton and Adri VAN AMELSVOORT (2001): «A practical approach to forensic earwitness identification: constructing a voice line-up». *Problems of Forensic Sciences*, 47(1), 238-241.

BROEDERS, Ton, Tina CAMBIER-LANGEVELD, and Jos VERMEULEN (2002): «Case report: arranging a voice lineup in a foreign language». *The International Journal of Speech, Language and the Law*, 9(1), 104-112.

BUTCHER, Andrew (1996): «Getting the voice line-up right: analysis of a multiple auditory confrontation», in Paul McCormack and Russell, Alison (eds.): *Proceedings of the 6th Australian International Conference on Speech Science and Technology*. Canberra: Australian Speech Science and Technology Association, 97-102.

DE JONG-LENDLE, Gea, Francis NOLAN, Kirsty MCDOUGALL, and Toby HUDSON (2015): «Voice lineups: a practical guide». Paper presented at the *18th International Congress of Phonetic Sciences*. Glasgow: IPA, 1-5.

DITTMAR, Norbert (1996): «Explorations in "Idiolects"», in Robin Sackmann and Budde, Monika (eds.): *Theoretical Linguistics and Grammatical Description: Papers in Honour of Hans-Heinrich Lieb*. Amsterdam: Benjamins,111-115.

GONZÁLEZ-RODRÍGUEZ, Joaquin (2014): «Evaluating automatic speaker recognition systems: an overview of the NIST speaker recognition evaluations (1996-2014) ». *Loquens*, 1(1), 1-15.

HOLLIEN, Harry (2002): *Forensic voice identification*. London: Academic Press.

HOLLIEN, Harry (2012): «On earwitness lineups». *Investigative Sciences Journal*, 4(1), 1-17.

HOLLIEN, Harry (2016): «The case for aural perceptual speaker identification». *Forensic Science International*, 269(1), 8-20.

KÖSTER, Olaf, Niels SCHILLER, and Hermann KÜNZEL (1995): «The influence of native-language background on speaker recognition», in Kjell Elenius and Branderud, Peter (eds.): *Proceedings of the 13th International Congress in Phonetic Sciences*. Stockholm: IPA, 306-309.

KÖSTER, Olaf and NIELS, Schiller (1997): «Different influences of the native language of a listener on speaker recognition». *Forensic Linguistics*, 4(1), 18-28.

LOAKES, Deborah (2006): «Variation in long-term fundamental frequency: measurements from vocalic segments in twins' speech», in Paul Warren and Watson, Catherine (eds.): *Proceedings of the 11th Australian International Conference on Speech Science & Technology*. Auckland, New Zealand: Australian Speech Science & Technology Association Inc., 205-210.

MANZANERO, Antonio Lucas and Susana BARÓN (2017): «Recognition and discrimination of unfamiliar male and female voices». *Behavior & Law Journal*, 3(1), 52-60.

NOLAN, Francis (1983): *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.

RASINGER, Sebastian ([2008] 2013): *Quantitative research in linguistics: an introduction*. London: Bloomsbury, 41-43.

RODRÍGUEZ BRAVO, Ángel, Patrícia LÁZARO PERNIAS, Norminanda MONTOYA VILAR, Josep Maria BLANCO, Dolors BERNADAS SUÑÉ, Daniel TENA PARERA, Ludovico LONGHI, and Josep Manel OLIVER COMES (2003): «Identificación perceptiva de locutores para la acústica forense: las RRV», in *II Congreso de la Sociedad Española de Acústica Forense*. Barcelona: SEAF, 23-34.

ROSE, Philip (2002): *Forensic speaker identification*. London: Taylor & Francis, 10-53.

SAN SEGUNDO, Eugenia (2014): «El entrenamiento musical y otros factores que pueden influir en el reconocimiento perceptivo de hablantes», in Yolanda Congosto (ed.): *Fonética Experimental, Educación Superior e Investigación*. Madrid: Arco Libros.

WEINREICH, Uriel, William LABOV, and Marvin HERZOG (1968): «Empirical foundations for a theory of language change», in Winfred Lehmann and Malkiel, Yakov (eds.): *Directions for Historical Linguistics*. Austin: University of Texas Press, 188-190.

TOMPKINSON, James and Dominic WATT (2018): «Assessing the abilities of phonetically untrained listeners». *Language and Law/ Linguagem e Direito*, 5(1), 19-37.

ZHANG, Cuiling and Tiejun TAN (2008): «Voice disguise and automatic speaker recognition». *Forensic Science International*, 175(2), 118-122.