

Recognition and discrimination of unfamiliar male and female voices

Reconocimiento y discriminación de voces desconocidas de hombre y mujer.

Antonio L. Manzanero¹

Susana Barón¹

(1) Universidad Complutense de Madrid, Madrid, España.

Email correspondencia: amanzane@ucm.es

Abstract

The aim of this study was to examine the ability to identify voices of unfamiliar people. In experiment 1, participants were tested in their ability to recognize the voice of unfamiliar man or woman in a target-present or target-absent earwitness identification line-up. Results showed that subjects generally matched 83.11% when the target voice was present and made 56.45% false alarms when it was missing. Discrimination was different from chance and subjects used liberal response criteria. In experiment 2, men and women tried to identify the same voices as in the previous experiment. Between stimulus presentation and the recognition task, subjects listened instrumental music for 2.38 minutes, with the purpose to hinder the possibility of the voice remaining active in the working memory. Results showed that the ability of men and women to identify an unfamiliar voice was null, in both cases with liberal response criterion. Men matched 12.06%, with 65.51% false alarms, and women 25.80% and 56.45% respectively. There were no differences in the ability to identify male and female voices, although women tend to indicate that the voice is present more than man even though no target voice was present.

Keywords: *memory, perception, identification, voice, earwitness testimony.*

Resumen

Se analizó la capacidad para identificar voces de personas desconocidas. En un primer experimento se presentaba una voz desconocida de un hombre o una mujer pronunciando una breve frase, para después tratar de discriminarla en una rueda de objetivo ausente o presente. Los resultados muestran que en general los sujetos acertaron un 83.11% de las veces en las que se presentó la voz objetivo y cometieron un 56.45% de falsas alarmas cuando no aparecía. La discriminación fue distinta del azar y los sujetos utilizaron criterios de respuesta liberales. En el segundo experimento entre la presentación del estímulo y la tarea de reconocimiento los sujetos escucharon una música instrumental durante 2.38 minutos, para dificultar que la voz presentada permaneciera activa en la memoria operativa. Los resultados mostraron que la capacidad para identificar una voz desconocida era nula, con un criterio de respuesta liberal en ambos casos. Los hombres acertaron un 12.06% de las ocasiones, con un 65.51% de falsas alarmas, y las mujeres un 25.80% y un 56.45% respectivamente. No se encontraron diferencias en la capacidad para identificar voces masculinas y femeninas, aunque las mujeres tienden a señalar más que los hombres aun cuando no esté presente la voz a identificar.

Palabras clave: *memoria, percepción, identificación, voces, testigos*

I. INTRODUCTION

Several studies (Yarmey, 2007) show how the ability to identify an unfamiliar voice heard only once for a short period of time, do not usually reach 50%. The most important factor in the ability to identify voices seems to be familiarity. Yarmey, Yarmey, Yarmey, and Parliament (2001) evaluated the role of familiarity in the ability to identify voices, considering four different levels of familiarity: the voices of people in the family, the voices that are moderately familiar from a working or a leisure environment, low familiarity voices, and unfamiliar voices. Results of their first experiment showed that 85% of the voices of familiar people were correctly identified, 5% were false alarms and 10% were omissions. Moderately familiar voices were correctly identified 79% of the times, 13% were false identifications and 8% were omissions. The low familiarity voices were identified 49%, 23% were false identifications and 28% were omissions. Unknown voices were correctly identified 55% and 45% were false identifications. Even the voices of relatives showed an error rate of 15%. In addition, the ability to identify unfamiliar voices is nearly 50%, significantly less than the ability to identify relatives or acquaintances.

On the other hand, a voice can be distinguished another by its different accent when pronouncing certain phonemes, for its tone (as it may be more acute or more severe), or pitch (to the extent that there are more melodic voices than others). Yarmey (1994) noted that one of the main factors that influence the accuracy in the discrimination of voices is the composition of the voice line-ups. The line-ups must be composed by several characteristics that must be taken into account: the speaking rate, its variability, monotony, expression, age, pronunciation, inflections, vibration, breaks and throaty.

Like in the description of a face, witnesses often provide poor information of what characterizes a particular voice. The ability to describe a voice is affected by different variables such as the amount of time that passes between the first encounter to the voice and the time of recognition. Yarmey (1991a) asked to describe a voice immediately, 24 hours later or a week later. Results showed that there was loss of information regarding the speaking rate, the variability of the voice, vibration, pauses, and throaty. Time delay did not seem to affect the information on expression, age, pronunciation or inflections, for less distinctive voices. However, the most distinctive voices did not seem to be

affected by these elapsed intervals.

The fact is that the voice is dynamic and it is also influenced by several variables such as emotion, fatigue, respiratory disorders, or other circumstances, such as if the person has an object in the mouth. On the other hand, the voice never reaches the listener directly but does so through some medium that can alter its characteristics; also environmental conditions generates resonances and interferences. However, several studies showed that the ability to identify a voice is not affected by the fact of listening through the phone (Kerstholt, Jansen, Van Amelsvoort, & Broeders, 2006; Yarmey, 2003).

Other factors involved are related to the duration of the vocal emission (Legee, Grosman, & Pieper, 1984), the frame time between the voice is heard and the recognition task (Kerstholt, Jansen, Van Amelsvoort, & Broeders, 2004, 2006; Yarmey & Matthys, 1992), the number of voices to identify and fillers (Carterette & Barnebey, 1975), its integration with facial information (Campanella & Belin, 2007), gender of sender and receiver (Bull & Clifford, 1984; McGehee, 1937; Wilding & Cook, 2000), distinctiveness of the voices (Yarmey, 1991b), language (Philippon, Cherryman, Bull, & Vrij, 2007) and stereotypes or beliefs of the witnesses (Yarmey, 1994). In addition, all these variables interact with each other. For example, in the case of emission duration and frame time, the latter is less affected the longer the duration of the emission is (Kerstholt et al., 2004, Yarmey & Matthys, 1992).

On the other hand, the ability to identify a voice does not seem to correlate with the ability of subjects to describe the characteristics of voices, nor the accuracy on describing the literal content or meaning of the messages (Yarmey, 2001). Thus, a witness may not be able to specify what was said but who said it, and viceversa.

Finally, other factors to consider would be the conditions of recovery and the task that is asked to the witness. For instance, different investigations (Kerstholt et al., 2006) show that when participants were asked to recognize a voice in a line-up in which no target voice was present, false identifications increased very significantly.

Regarding the difference between men and women to identify voices, McGehee (1937) found that men were more effective on identifying women's voices, while the voices of men were equally recognized by both genders. However, the literature shows that these data is not always confirmed. Bull and Clifford (1984) found that women have a better ability

to identify voices than men, and women's voices would be easier to distinguish than those of men. Wilding and Cook (2000) found that men distinguish equally well the voices of men and women, while women distinguished the voices of women better than men's voices. What causes these differences? Perhaps evolutionary aspects may be responsible for a neurological specialization on the discrimination between male and female voices for men and women; as Sokhi, Hunter, Wilkinson, and Woodruff (2005) found male and female voices activate different brain areas during gender attribution, using functional magnetic resonance imaging (fMRI).

Therefore, the aim of this study was to analyze the ability to discriminate unfamiliar voices of people saying short phrase immediately before, on target-absent or target-present line-ups (experiment 1). In this procedure, the ability of men and women to identify male and female voices in target-present and target-absent line-ups was also analyzed (experiment 2), given the limited but varied results regarding gender differences. In the first experiment, it was analyzed subjects' ability to discriminate voices (and therefore would be primarily a perceptual task), while in the second experiment, it was assessed if the participants were able to identify the voices (it would be a memory task, with a short retention interval). Thus, we try to answer the general question of what differences exist in the ability to discriminate and identify unfamiliar male and female voices. As the discrimination task is easier than the identification task, it is hypothesized that subjects would be able to perform both tasks above chance, but they will have more difficulty performing the identification task. No difference is expected on the basis of gender.

II. EXPERIMENT 1

A. Participants

54 subjects (50 women), with a mean age of 21.35 years ($SD = 4.28$), participated in this experiment. All were students of the Faculty of Psychology at the Universidad Complutense de Madrid, and collaborated in the study voluntarily and without compensation.

B. Procedure

In the present experiment generic capability to discriminate voices of men and women in target-present or target-absent line-ups was assessed. Each subject participated in 24 trials in which they heard an unfamiliar voice (12 trials for male and 12 for female voices). Immediately after, the participants tried to identify the voice presented. Each line-up was composed of five voices, also from unfamiliar people who uttered the same sentence.

Throughout the experiment, subjects listened twelve different voices, six female and six male voices. Each voice was the target twice, one for target-present and other for target-absent line-up. The voices were presented sequentially and randomly.

Participants had to answer whether the original voice was in the line-up or absent. If the voice was in the line-up, subjects had to indicate the sequence number occupied by the voice which they believed corresponded to the target. 33 subjects participated in trials with female voices and 21 subjects participated in trials with male voices.

C. Material

The samples were composed of six men and six women's voices. All of them uttered the phrase "The capital city of Italy is Rome." Average duration for male voices was 1.69 sec. ($SD = 0.20$) and 1.83 sec. ($SD = 0.21$) for female voices. Thus, in this experiment we try to reproduce the actual conditions present in most short duration crimes such as robberies that happen in less than a minute where the aggressors speak in a short-phrases mode similar to "Give me your money". However, emotional factors were set aside since previous research (e.i., Schirmer, Zysset, Kotz, & von Cramon, 2004) has shown that these factors differentially affect men and women.

Voices were randomly selected among students of first courses of the Faculty of Psychology at the Universidad Complutense de Madrid, in an age range from 19 to 21 years. All students were from the same Spanish region (Madrid) in order to prevent any specific accent or any element of distinction. All voices had similar characteristics (see Figure 1 for male voices and Figure 2 for female voices). The voices were recorded in digital form.

Figure 1. Acoustic spectrum for male voices.

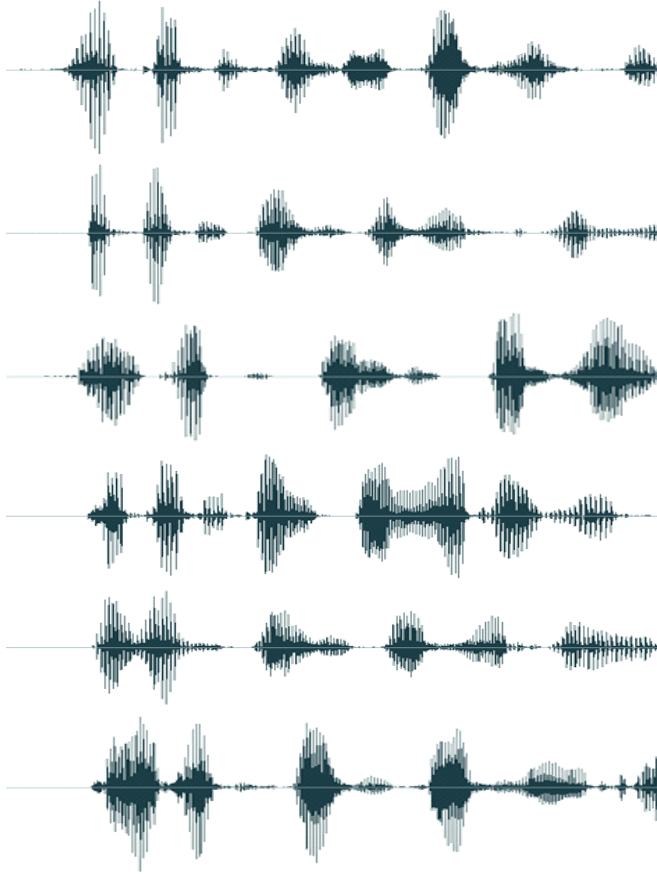
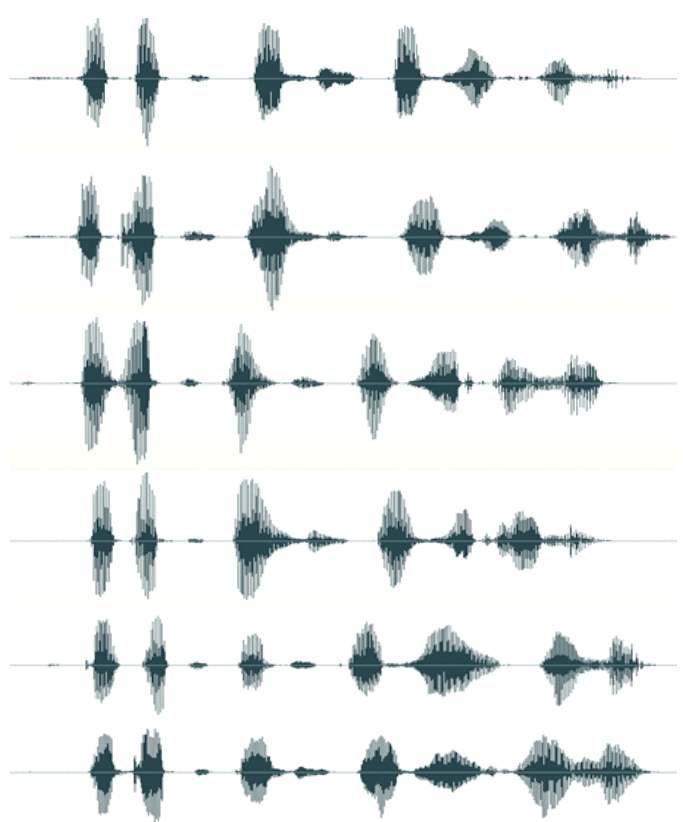


Figure 2. Acoustic spectrum for female voices.



D. Results and discussion

As shown in Table 1, data show that the subjects matched 83.11% when the target voice was present and 56.45% false alarms when the target voice was gone. The participants were able to discriminate above chance ($d'=.796$, $SD=.078$, $Zd'=10.232$, $p<.05$). Participants used a liberal criterion ($c=-.561$, $SD=.039$), when making decisions about the presence of target voices in the discrimination task, i.e. significantly tended to assert that the target voice was present even when it was not there ($Zc=14.409$, $p<.05$).

Factorial analysis (ANOVA) showed that male voices were better recognized ($M_{hit} = 10.67$, $SD = 0.79$, $F(1,53) = 21.786$, $p<.001$, $\eta^2 = .295$; $M_{false\ alarm} = 5.38$, $SD = 2.37$, $F(1,53) = 15.091$, $p<.001$, $\eta^2 = .225$) in comparison with female voices ($M_{hit} = 8.73$, $SD = 1.79$; $M_{false\ alarm} = 7.58$, $SD = 1.77$).

Both male and female voices were discriminable above chance, although the first voices are discriminated better than the latter ($M_{d'} = .901$, $SD = .164$, $Z_{d'} = 5.480$, $p<.05$, confidence interval = [.581, 1.221], Cohen's $d = 7.329$, effect size $r = .964$). The response criterion was liberal in both cases, though more in the case of female voices ($M_c = -.565$, $SD=.082$, $Z_c = 6.873$, $p<.05$, confidence interval = [-.725, -.405], Cohen's $d = .275$, effect size $r = .136$).

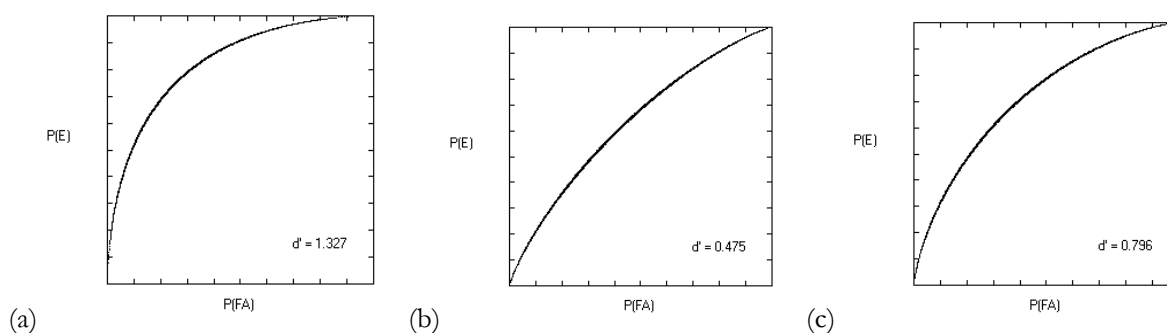
From these results, the question was if men and women were equally able to identify male and female voices. It was performed the following experiment in which, unlike the previous study, subjects did not perform the task immediately, but after a short retention interval during which they listen to instrumental music.

Table 1. Scores and percentages (in parentheses) for each type of response condition.

	Voices		
	Male	Female	Global
Hit	224 (44.88)	288 (37.89)	512 (40.66)
Omission	28 (5.61)	76 (10.00)	104 (8.26)
False Alarm	113 (22.64)	250 (32.89)	363 (28.83)
Correct Rejection	134 (26.85)	146 (18.92)	280 (22.23)
d'	1.327* ($SD=.132$)	.475* ($SD=.098$)	.796* ($SD=.078$)
c	-.557* ($SD=.066$)	-.573* ($SD=.049$)	-.561* ($SD=.039$)

* Significant $p < .05$

Figure 3. Graphic representations of the ROC curve for (a) male voices, (b) female voices and (c) total.



III. EXPERIMENT 2

A. Participants

In this experiment, 60 participants (31 women and 29 men; $M_{age} = 28.03$ and 27.06 years, $SD = 8.30$ and $SD = 6.61$, respectively) tried to identify the voice of a man and a woman previously heard while delivering a brief sentence. Participants collaborated in the study voluntarily and without compensation.

B. Procedure

Each subject participated in two trials, one for male and one for female voice (counter-balanced). Once participants heard the target voice, they listened to instrumental music for 2.38 minutes. Then half of the participants should identify whether the target voice was present in target-present line-ups, and the other half in target-absent line-ups. The line-ups were made of five voices. The voices were presented sequentially and randomly. Participants performed the test individually.

C. Material

The voices used in this experiment were the same as in

Experiment 1. In order to hinder participants to maintain activated the voices previously presented in the working memory, they heard during the retention interval an instrumental theme, "Erinnerung" performed by Klaus Doldinger (1981), the soundtrack of the movie "Das Boot", with duration of 2.38 minutes.

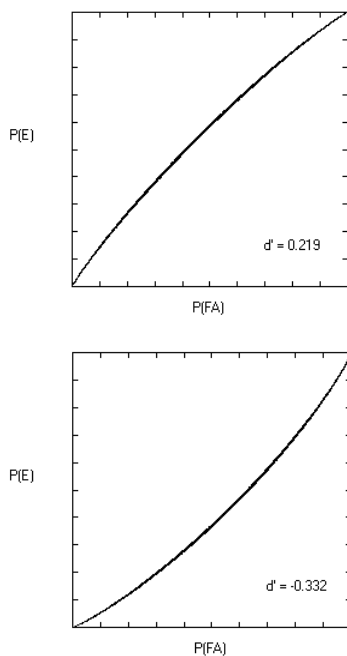
D. Results and discussion

1) Identification of male and female voices

For male voices, when the target voice was present in the line-up, participants matched only 40%; when it was absent there was a 60% false alarms. For female voices, the hits were 36.66%, when the target voice was present; and 80% false alarms when the target voice was absent (see Table 2).

For both types of voices, discriminability index (d') did not differ significantly from chance responses, which means that the ability of participants to identify male and female voices after a short retention interval is null ($M_{d'} = .219$, $SD = .420$, $Z_{d'} = .521$, $p = n.s.$, $M_{d'} = -.332$, $SD = .443$, $Z_{d'} = -.750$, $p = n.s.$, respectively).

Figure 4. Graphic representations of the ROC curve for (a) male voices, and (b) female voices.



In both cases, the response criterion was liberal both for female voices ($M_c = -.958, SD = .221, Z_c = -4.325, p < .05$, confidence interval = [-1.389, -.526]) and for male voices ($M_c = -.732, SD = .210, Z_c = -3.488, p < .05$, confidence interval = [-1.141, -.0323]). ROC graphics were inserted below (Figure 4).

2) Ability of men and women to identify voices

As shown in Table 2, when the target voice is present in the line-up, men are less right (23.33%), but they also make fewer false alarms when the voice is not present (63.33%), compared with women (53.33% and 76.66%, respectively).

However, when comparing the ability to discriminate unfamiliar voices in men ($M_d = .008, SD = .506, Z_d = .015, p = n.s.$) and women ($M_d = -.126, SD = .394, Z_d = -.319, p = n.s.$) no difference was found between the two groups ($M_d = -.059, SD = .641, Z_d = -.092, p = n.s., Cohen's d = .295$, effect size $r = .146$). In both cases the ability to discriminate was null, i.e. no different from chance (see Table 2).

Table 2. Scores and percentages (in parentheses) for each type of response condition.

		Voices				Total
		Male		Female		
		Target present	Target absent	Target present	Target absent	
Men N = 29	Hit	5 (35.71)	-	2 (14.28)	-	7 (12.06)
	Omission	1 (7.14)	-	1 (7.14)	-	2 (3.44)
	False Alarm	8 (57.14)	10 (66.66)	11 (78.57)	9 (60.00)	38 (65.51)
	Correct Rejection	-	5 (33.33)	-	6 (40.00)	11 (18.96)
	d'	.186 (SD = .675)		-.306 (SD=.796)		.008 (SD = .506)
	c	-.874 (SD = .338)*		-.583 (SD=.398)		-.761 (SD = .253)*
Women N = 31	Hit	7 (43.75)	-	9 (56.25)	-	16 (25.80)
	Omission	2 (12.50)	-	2 (12.50)	-	4 (6.45)
	False Alarm	7 (43.75)	8 (53.33)	5 (31.25)	15 (100.00)	35 (56.45)
	Correct Rejection	-	7 (46.66)	-	0 (.00)	7 (1.29)
	d'	.292 (SD = .42)		-1.593 (SD = .525)*		-.126 (SD = .394)
	c	-.619 (SD = .271)*		-1.705 (SD = .262)*		-.904 (SD = .197)*

* Significant $p < .05$

Regarding the response criteria it was observed that both groups tended significantly to liberal criteria ($M_c = -.761, SD = .253, Z_c = -3.007, p < .05$, confidence interval = [-1.254, -.268], for men; and $M_c = -.904, SD = .197, Z_c = -4.594, p < .05$, confidence interval = [-1.288, -.521], for women); that is to say, there was a tendency to point out one of the voices

presented in the test phase, even if the voice did not match the target. Women used significantly more liberal criteria than men ($M_c = .833, SD = .321, Z_c = -2.596, p < .05, Cohen's d = .630$, effect size $r = .300$). ROC graphics were inserted below (Figure 5).

Figure 5. Graphic representations of the ROC curve for (a) men, and (b) women.

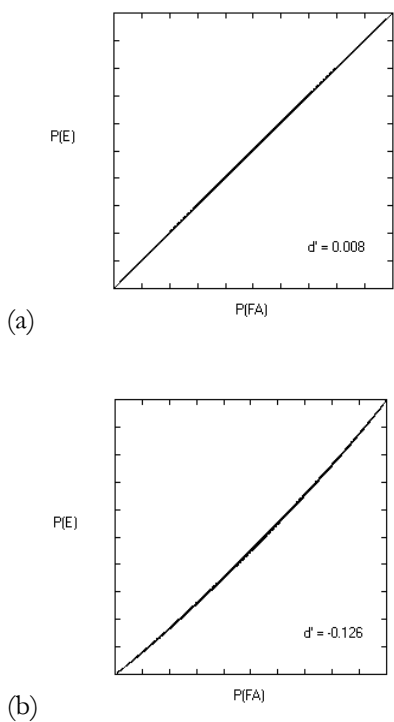
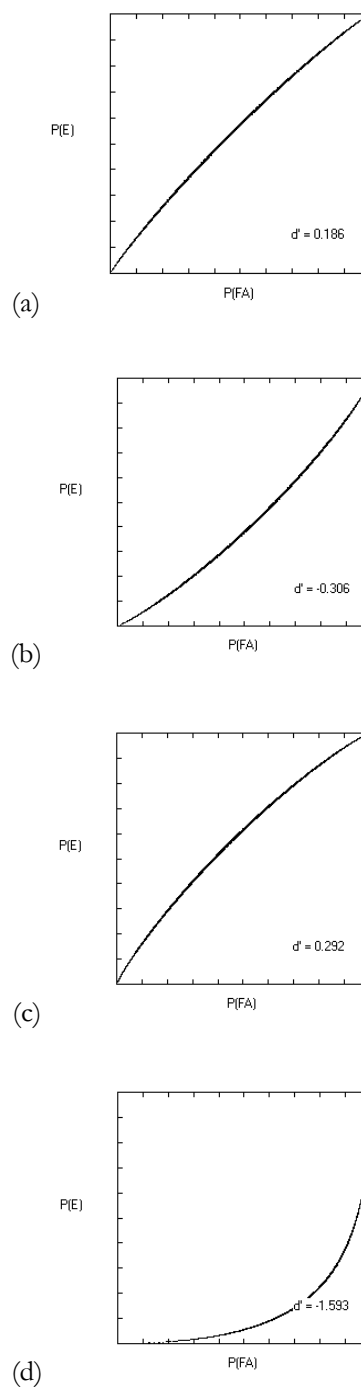


Figure 6. Graphic representations of the ROC curve for (a) men and male voices, (b) men and female voices, (c) women and male voices, and (d) women and female voices.



3) Ability of men and women to identify voices of men and women

Significance tests on discriminability (d') and response criteria (c) in men showed that the two indexes did not differ significantly depending on the type of voice to identify, $Z_{d'} = -.057$, $p = \text{n.s.}$; $Z_c = -1.395$, $p = \text{n.s.}$ For women, the significance analysis showed no differences in the ability to discriminate, depending on the type of voice to identify, $Z_{d'} = -1.200$, $p = \text{n.s.}$, although they did differ significantly in response criteria, $Z_c = -4.288$, $p < .05$, since they were more liberal in trying to identify a female voice ($M_c = -1.705$) than in trying to identify a male voice ($M_c = -.619$). Considering the four conditions, we found no significant differences in the ability to discriminate ($M_{d'} = -.355$, $SD = .392$, $Z_{d'} = -.906$, $p = \text{n.s.}$), but significant differences in response criteria ($M_c = -.945$, $SD = .196$, $Z_c = -4.820$, $p < .05$).

The only condition in which the discriminability index (d') was significantly different from chance was in the ability of women to identify voices of other women. The observed data showed that this capacity is significantly below chance (see Table 2 and Figure 6). When the target female voice was not in the line-up, the women pointed to another voice in 100% of cases.

IV. CONCLUSIONS

As pointed out in previous researches (Yarmey, 2007), the results of the experiments described here showed that even when discrimination of an unfamiliar voice heard during a short period of time is possible, the ability to identify it after a short retention interval is null.

In a high percentage of cases, participants tend to point erroneously at one of the voices, whether the target voice was in the line-up or not, in line with the data found by Kerstholt et al. (2006). This lack of ability to identify is in part due to the tendency of subjects to point at one of the voices regardless of whether or not the target voice was in the line-up. In these regard, there is no variance whether the listener is men or woman. And the gender of the sample provides no contrast either. On the contrary, it stands out that when it comes to women trying to identify a woman's voice, false alarms reach a rate of 100% when the line-up does not show the target voice.

From an applied point of view to a judicial environment, it must be warned that here is a high possibility that witnesses will falsely identify the voice of a perpetrator when asked to identify the voice in a line-up of voices, on

the grounds that there is a common inability to identify unfamiliar voices, even below chance in some conditions. Moreover, in real scenarios his capability could be even lower if we consider that in general the line-up can be carried out after a longer period of time (Kerstholt et al., 2004, 2006; Yarmey & Matthys, 1992), and the ability to process the voice features will be affected negatively by the anxiety of witnesses and victims of crimes that could hamper the discriminability of the target voice. It must not be forgotten that the laboratory conditions in which participants were in the previous experiments were suitable (hearing was optimal, there were no distracting factors, in Experiment 1 the voices were sufficiently discriminable, the line-up was conducted shortly after perceiving the voices, and the subjects knew in advance that the aim was to identify the voice). By contrast, emissions of longer duration may be more easily identified (Legge, Grosmann, & Pieper, 1984).

V. ACKNOWLEDGEMENTS

This work was made possible through the collaboration of Carla Paola Russo and Desiree Fretes involved in the recollection of voice samples and administration of the tests in experiment 2, and Rocío Vallet who participated in recording of date for experiment 1. The research is part of False Guilty Project, developed at the Universidad Complutense de Madrid (<http://falsoculpable.blogspot.com>), in order to analyze the main causes leading to an innocent accused of crimes due to a false identification.

VI. REFERENCES

- Bull, R., & Clifford, B. R. (1984). Earwitness voice recognition accuracy. In G. L. Wells & E. F. Loftus (Eds.), *Eye witness testimony: Psychological perspectives* (pp. 92-123). Cambridge: Cambridge University Press.
- Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Sciences*, 11(12), 535-543.
- Carterette, E. C., & Barneby, A. (1975). Recognition memory for voices. In E. Cohen y G. Nottebohn (Eds.), *Structure and processes in speech perception* (pp. 246-265). New York: Springer.
- Kerstholt, J. H., Jansen, N. J. M., Van Amelsvoort, A. G., & Breeders, A. P. A. (2004). Earwitnesses: Effects of speech duration, retention interval and acoustic environment. *Applied Cognitive Psychology*, 18, 327-336.
- Kerstholt, J. H., Jansen, N. J. M., Van Amelsvoort, A. G., & Breeders, A. P. A. (2006). Earwitnesses: Effects of accent, retention and telephone. *Applied Cognitive Psychology*, 20, 187-197.
- Legge, G. E., Grosmann, C., & Pieper, C. M. (1984). Learning unfamiliar voices. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 298-303.
- McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology*, 27, 249-271.
- Philippon, A., Cherryman, J., Bull, R., & Vrij, A. (2007). Lay people's and police officers' attitudes towards the usefulness of perpetrator voice identification. *Applied Cognitive Psychology*, 21, 103-115.
- Schirmer, A., Zysset, S., Kotz, S. A., & von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *Neuroimage*, 21(3), 1114-1123.
- Sokhi, D. S., Hunter, M. D., Wilkinson, I. D., & Woodruff, P. W. R. (2005). Male and female voices activate distinct regions in the male brain. *Neuroimage*, 27, 572-578.

- Manzanero, A. L. & Barón, S. (2017). Recognition and discrimination of unfamiliar male and female voices. *Behavior & Law Journal*, 3(1), 52-60.
- Wilding, J., & Cook, S. (2000). Sex differences and individual consistency in voice identification. *Perceptual and Motor Skills*, 91, 535–538.
- Yarmey, A. D. (1991a). Voice identification over the telephone. *Journal of Applied Social Psychology*, 21, 1868–1876.
- Yarmey, A. D. (1991b). Descriptions of distinctive and non-distinctive voices over time. *Journal of the Forensic Science Society*, 31, 421–428.
- Yarmey, A. D. (1994). Earwitness evidence: Memory for a perpetrator's voice. In D. F. Ross, J. D. Read & M.P. Tolia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 101-124). New York: Cambridge University Press.
- Yarmey, A. D. (2001). Earwitness descriptions and speaker identification. *Forensic Linguistics*, 8, 113–122.
- Yarmey, A. D. (2003). Earwitness identification over the telephone and in field settings. *The International Journal of Speech, Language and the Law*, 10, 62-74.
- Yarmey, A. D. (2007). The psychology of speaker identification and earwitness memory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Tolia (Eds.), *The handbook of eyewitness psychology, Vol II: Memory for people* (pp. 101-136). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Yarmey, A. D., & Matthys, E. (1992). Voice identification of an abductor. *Applied Cognitive Psychology*, 6, 367–377.
- Yarmey, A. D., Yarmey, A. L., Yarmey, M. J., & Parliament, L. (2001). Commonsense beliefs and the identification of familiar voices. *Applied Cognitive Psychology*, 15, 183-299.