ISSN: 1889-1861

THE EUROPEAN JOURNAL OF PSYCHOLOGY APPLIED TO LEGAL CONTEXT

Volume 2, Number 2, July 2010

The official Journal of the
SOCIEDAD ESPAÑOLA DE PSICOLOGÍA JURÍDICA Y FORENSE
Website: http://www.usc.es/sepjf

Official Journal of the *Sociedad Española de Psicología Jurídica y Forense*
(www.usc.es/sepjf)
Published By: SEPJF.
Volume 2, Number, 2.
Order Form: see www.usc.es/sepjf
Frequency: 2 issues per year.
ISSN: 1889-1861.

The European Journal of Psychology Applied to Legal Context, 2010, 2(2)
www.usc.es/sepjf

CONTENTS

Articles

Judicial judgement-making and legal criteria of testimonial credibility
Mercedes Novo and Dolores Seijo 91

Juvenile delinquency and young offender: bibliographical and
bibliometric review of two perspectives of study
Francisco Javier Rodríguez, Luis Rodríguez-Franco,
Javier López-Cepero, and Carolina Bringas 117

Modelling alcohol consumption during adolescence
using zero inflated negative binomial and decision trees
Elena Gervilla, Berta Cajal, Joan Roca, and Alfonso Palmer 145

Mobile phone quality vs. Direct quality:
How the presentation format affects earwitness identification accuracy
Lisa Öhman, Anders Eriksson, and Pär Anders Granhag 161

A stepwise approach to identify intellectual disabilities
in the criminal justice system
Erik Søndenaa, Tom Palmstierna, and Valentina Cabral Iversen 183

The European Journal of Psychology Applied to Legal Context, 2010, 2(2): 161-182
www.usc.es/sepjf

MOBILE PHONE QUALITY VS. DIRECT QUALITY: HOW THE PRESENTATION FORMAT AFFECTS EARWITNESS IDENTIFICATION ACCURACY

Lisa Öhman*, Anders Eriksson**, Pär Anders Granhag*

* Department of Psychology, University of Gothenburg, Sweden
** Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Sweden

(Received 30 November 2009; revised 13 April 2010; accepted 19 April 2010)

Abstract

The present study aimed to gain insight into the effect of mobile phone quality on voice identification using an ecologically realistic design. A total of 165 participants were exposed to an unfamiliar voice, either directly recorded or mobile phone recorded, for 40 seconds. After a two week delay, they were asked to identify the target-voice in a 7 voice target-present line-up. We used a between-subjects design, where half of the subjects were exposed to a directly recorded line-up, and the other half to a mobile phone recorded line-up. Data analysis did not show any significant effect of presentation format or line-up format. These results suggest that the detrimental effect on voice recognition suggested by the poorer sound quality of mobile phone recordings is minimal. They also indicate that there is no benefit from conducting a mobile phone recorded line-up, if the voice is originally heard over a mobile phone. More research is needed, however, before definitive conclusions may be drawn. The overall accuracy for correct identifications was 12.7% which is expected by chance. Further, one particular foil attracted 54% of all false identifications. Future research should focus on explaining why earwitnesses perform so poorly and develop methods to improve identification accuracy.

Keywords: Earwitnesses, mobile phone, voice identification, line-up, presentation format.

Resumen

Por medio de un diseño ecológicamente realista, nos planteamos un estudio con el objeto de profundizar en el efecto de la calidad del teléfono móvil en la identificación de la voz. Un total de 165 participantes fueron expuestos durante 40 segundos a una voz no familiar, bien grabada directamente, bien en teléfono móvil. Tras un intervalo de demora de dos semanas, se les pidió que identificaran la voz en una rueda con siete voces. Con un diseño inter-sujetos en el que la mitad de los participantes fueron expuestos a una rueda de identificación grabada directamente, y la otra mitad a una rueda de identificación grabada con un teléfono móvil. El análisis de datos no mostró efectos significativos mediados por el formato de presentación o e la rueda. Estos resultados sugieren que el efecto perjudicial en el reconocimiento de la voz derivado de la baja calidad de sonido de las grabaciones de los teléfonos móviles es mínimo. Asimismo, también soportan que, si la voz es originariamente, escuchada a través de un teléfono móvil, no hay inconveniente alguno en llevar a cabo una rueda grabada con teléfono móvil. No obstante, es necesaria más investigación antes de establecer conclusiones definitivas. La tasa de exactitud de identificaciones correctas estaba en el 12,7%, lo que es esperado por azar. Además, un distractor atrajo el 54% de las falsas identificaciones. La investigación futura debería centrarse en explicar porqué los testigos auditivos ejecutan tan pobremente y en desarrollar métodos para mejorar la exactitud de la identificación.

Palabras clave: Testimonio auditivo, teléfono móvil, identificación de voz, rueda de identificación, formato de presentación.

Correspondence: Lisa Öhman, Department of Psychology, University of Gothenburg, P.O. Box 500, 405 30 Gothenburg, Sweden. E-mail: lisa.ohman@psy.gu.se

Introduction

The memory of a perpetrator's voice by a witness or a victim may be key evidence and play an important role for crimes committed under conditions of darkness, disguised perpetrators, obscene phone calls, ransom demands and other threatening phone calls. However, earwitness identification has not nearly been investigated to the same extent as eyewitness identification (Deffenbacher et al., 1989, Yarmey, 1995).

It is quite common that criminal cases involve voices mediated by telephone. Clifford, Bull, and Rathborn (1980) note that it is "evident from our research ... into past criminal cases involving voice identification that a vast number of such cases involve the use of the telephone" (p. 280). Hirson, French, and Howard (1995) report that in "well over 90 per cent of cases, speaker identification for forensic purposes involves the comparison of a disputed speech sample recorded off a telephone line with reference speech samples elicited in face-to-face interviews with suspects" (p. 230).

As is well known, telephone transmission influences the sound quality negatively but the effect on speaker identification has been examined in only a few studies. In early studies the limited bandwidth of telephones was simulated by bandpass filtering of direct recorded speech (e.g. Rothman 1977; 1979). Later studies used speech recorded over land line phones (e.g. Rathborn, Bull, & Clifford, 1981). In recent years, the widespread use of mobile phones has increased the number of crimes where mobile phones are used. Jonas Lindh (p.c.) from Voxalys (a Swedish company that performs forensic analyses for the Swedish police) estimates that around two thirds of the cases involve mobile phone calls. Peter French (p.c.) at the Forensic Speech and Acoustics Laboratory in York, U.K. reports that "The number of forensic recordings of calls made from mobiles has not only increased in proportion to the increased number of mobiles relative to landlines, but because many criminals use unregistered mobiles for their communications in the belief that their calls cannot be traced back to them", a trend confirmed by Michael Jessen (p.c.) at the German Federal Criminal Police Office (BKA) who reports that "the involvement of mobile phones in crimes has increased over the years".

Mobile phone transmission affects sound quality in partly the same way as land line phones do (e.g. limited bandwidth, transmission losses and the effect of usually poor microphone quality) but not necessarily in identical ways and also in other ways. The digital coding is, however, likely to be responsible for most of the quality difference. There are two principle reasons for this. The transmission speed is greater for landline phones (typically 64 kb/s vs. 24.40 kb/s or less). The speed of the wireless transmission between the mobile phone and the mobile network varies between 4.75 kb/s and 24.40 kb/s and may change many times during a single call. "From the standpoint of a medium through which information is transferred, such channels can be significantly poorer than their landline counterpart" (Guillemin & Watson, 2008, p. 197). In our own recordings we have, for example, observed differences in measured formant values for the first two formants on the order of 150 Hz on average between mobile and direct recordings. We are not aware of any studies where the effect of these differences for aural voice recognition or memory has been studied, but it has been shown to have a considerable effect in automatic speaker recognition (Brümmer & Strasheim, 2009).

There are also considerable differences between mobile calls and landline calls with respect to speaking style. For example it has been found that people have a tendency to speak louder when using a mobile phone, and because of mobility there is often more background noise where mobile phones are used (Byrne & Foulkes, 2004; Foulkes & Barron, 2000).

It is therefore necessary to study the effect of the use of mobile phones in more detail. Of the few studies that have been carried out involving telephones, only one study has, as far as we are aware, used mobile phones (Kerstholt, Jansen, Van Amelsvoort & Broeders, 2006)

A view sometimes expressed is that when a voice is initially heard over the telephone it is desirable to conduct the line-up using voices recorded over the telephone. Rathborn et al. (1981) note that "when a ... crime ... involves the identification of a voice heard initially over the telephone, it is common practice for the police to request the suspect to telephone the victim and repeat the same phrases ... This procedure suggests that the police share the commonsense belief that voice recognition will be enhanced if the test takes place under the same circumstances as the initial hearing" (p. 280). However, previous research does not support that belief. Rather, studies have shown that using a telephone recorded line-up when the voice is

originally heard over a telephone will not result in greater accuracy (Kerstholt et al., 2006; Rathborn et al., 1981). If it can be shown that telephone recorded line-ups do not improve earwitness accuracy in those cases, it would, to some degree, facilitate the work of the police and phonetic experts since direct recorded voice samples may then be used for the line-ups without causing any negative effects on recognition.

Hence, the focus of the present study is on the effect of the presentation format (mobile phone vs. direct recording) for voice recognition and investigating the usefulness of conducting a line-up with mobile phone recorded voices if the target voice is originally heard over a mobile phone. This will be tested with an experimental set-up aiming for a high degree of ecological validity. As our review below will show only a handful of comparable studies exist, leaving many unanswered questions.

Previous research

For technical reason, as was explained above and confirmed by our daily experience, the sound quality mediated by telephone, and mobile phones in particular, is vastly inferior to that of a direct recording of good quality.

The effect of the degraded quality on the acoustic analysis of speech for forensic purposes has been examined in some studies (Byrne & Foulkes, 2004; Künzel, 2001). In these studies the limited telephone bandwidth has been shown to affect the position of the lowest resonance (first formant) in vowels which may affect the reliability of acoustic speaker comparison.

The available studies concerning the effect of telephones on voice identification are few and they vary considerably with respect to methodology. Rothman (1977) let listeners hear six paired voice samples, using the original direct recordings or bandpass filtered versions simulating telephone quality. The task was to decide if the speaker was the same in the two samples. He found that the simulated telephone quality voices were more difficult to identify than the voices in the original recording. In a forensically slightly more relevant study, Rathborn et al. (1981) used direct recordings and telephone recordings. The participants heard either a direct recorded voice or a telephone recorded voice and were confronted with a six-voice target-present line-up consisting of direct recorded or telephone recorded voices immediately after hearing the target voice. It was found that

subjects who were presented with a direct recorded target and a direct recorded line-up performed significantly better than the subjects in the other three conditions. There were no significant differences between the conditions where telephone recordings were used. These studies imply that telephone quality speech decreases recognition performance in general. However, the use of pairwise comparison (Rothman, 1977) or several line-ups per subject (Rathborn et al., 1981) is not comparable to conditions in a real life investigation. In studies were subjects have been presented with a single target voice and a subsequent line-up (with varying delay) an effect of telephone quality has not been found. Yarmey (2003) compared speaker identification accuracy between subjects who spoke to the target over the telephone with subjects who spoke to the target face-to-face. He found no significant difference in identification accuracy between the two groups. Neither did Perfect, Hunt, and Harris (2002) in their study. Although they did not find a difference in identification accuracy, they found that subjects who had heard a voice over the telephone reported lower confidence compared to those who had heard a direct recorded voice. The Kerstholt, et al. (2006) study used a mobile phone, but the set-up was quite unusual in that the speakers did not use the mobile phone directly. Instead recorded speech was presented over a loudspeaker and a mobile phone held close to the loudspeaker transmitted the sound to an ISDN centre. However, they did not find any significant difference in identification accuracy between the subjects in the telephone condition and those who heard a direct recorded voice. The results in these three studies thus imply that telephone quality does not have an effect on voice identification accuracy, in contrast to the results obtained by Rathborn et al. (1981) and by Rothman (1977).

This brief review shows that the few previous studies involving telephones have shown mixed results. In addition, only one previous study has tested the effect of using the now most common type of telephone, the mobile phone. This strongly suggests that more research is needed in his area.

Retention interval

In field situations there is often a time delay of one week or more between the crime and an opportunity for a voice identification line-up. In experimental studies it is not uncommon, however, to use much shorter retention intervals like 24 hours or less (e.g. Philippon, Cherryman, Bull, & Vrij, 2007; Saslove & Yarmey,

1980, Yarmey, 2001). To ensure ecological validity a 24 hour retention interval is, in our opinion, an absolute minimum, while longer intervals, on the order of one week or two, are to be preferred for reasons of realism. Studies using longer retention intervals show mixed results. Some studies have found no difference in performance between a 1 week and a 2 week delay (Van Wallendael, Surace, Hall-Parsons, & Brown, 1994), between a 1 week and an 8 week delay (Kerstholt et al., 2006) or between a 2 week and an 8 week delay (McGehee, 1944). Other studies, however, have found a significant drop in performance between week 2 and week 3 (Clifford & Denot cited in Bull & Clifford, 1984) and also that the false alarm rate increases after 7 days (Yarmey & Matthys, 1992). The only study, to our knowledge, that has tested the effect of retention time on recognition of voices heard over the telephone, showed no difference in accuracy between a 2-hour delay and a delay of 2–3 days (Yarmey, 1991). In order to obtain a reasonable degree of ecological validity we chose a retention interval of 2 weeks for the present study.

The Present Study

The aim of the present study was to investigate the influence of presentation and line-up format (direct recording vs. mobile phone) on voice recognition accuracy. Based on suggestions in previous research and the considerable technical differences resulting in corresponding sound quality differences, we hypothesized that subjects who have heard a direct recorded voice would perform better at voice recognition than subjects who have heard a voice recorded through a mobile phone (Hypothesis 1). In line with previous studies, we also hypothesized that a mobile phone recorded line-up would not improve voice recognition accuracy in cases where the target was recorded via a mobile phone (Hypothesis 2).

It is a well known fact that there is a strong tendency among legal fact finders to rely heavily on witnesses' confidence when inferring the accuracy of a statement or an identification (e.g., Cutler, Penrod, & Stuve, 1988). To date, relatively little is known on the realism of earwitnesses' meta-cognitive assessment, but the few studies on this topic have found that the confidence judgments made by earwitnesses should be assigned little diagnostic value (e.g. Olsson, Juslin, & Winman, 1998; Yarmey, 2001; Öhman, Eriksson, & Granhag, unpublished). Hence, we decided to

explore this issue further, and based on past research we expected no or only a weak relationship between accuracy and confidence (Hypothesis 3).

Method

Participants

A total of 174 individuals participated in the study. Since linguistic properties of the speech material may have an influence on both memory for the mock perpetrators' voice and similarity judgements in the line-ups it was a requirement that participants must be native speakers of Swedish or at least have native like fluency in the language. For social reasons we chose not to discriminate any subject from participating in the test, but in the analyses the results from those who did not meet the linguistic requirements were eliminated. Altogether 9 of the 174 participants were excluded. The remaining group of 165 subjects (*M*(age) = 24.7, *S D*= 7.1) consisted of 112 women and 53 men. The subjects were individuals who had announced their interest to participate in studies at the University of Gothenburg. Each participant was given a compensation equivalent of 100 SEK (approximately 14 USD) for participating.

Design

A 2 (Presentation format: direct vs. mobile phone) x 2 (Line-up presentation: direct vs. mobile phone) between-group design was employed. Each participant was randomly assigned to one of the four conditions.

Recording the speech material

Speech material for the study was selected from a database of 25 recordings of speakers reading the mock incriminating call from a manuscript and describing a walk through Gothenburg city, guided by a set of pictures, in an informal interview situation. Each recording took 10–15 minutes. Most of the recordings were made at the department of linguistics at University of Gothenburg but also, in some cases, in the offices of a company where we recruited speakers. All speakers were from the Gothenburg area and spoke an educated form of the regional dialect. The chronological age of the speakers ranged from 25 years of age to 52.

To safeguard ecological validity we preferred good quality recordings rather than "studio quality" for the direct recordings. All recordings were made using a medium priced solid state recorder (Zoom, H2) with a built-in microphone. The recorder was placed on a table about 60 cm in front of the speaker. Each session was simultaneously recorded via a mobile phone calling a land line phone connected to a computer in an office at the Department of Linguistics, Gothenburg University to obtain identical telephone quality recordings. The mobile phone calls were recorded digitally at 16 kHz/16 bit (wav). The direct recordings were originally sampled at 44.1 kHz/16 bit (wav) but later re-sampled to 16 kHz/16 bit to match the mobile phone recordings. All recordings used in the experiments were thus 16 kHz/16 bit wav files.

Selecting and preparing the recordings for the line-ups

It turned out that some of the recordings had to be discarded due to technical flaws (mobile phone interference, audibly different room echo or insufficient content). From the remaining 16 recordings, the "perpetrator" was then chosen by listening to all recordings and selecting one that sounded reasonably "involved" and not "read".

Foils for the line-ups were selected using a perceptual evaluation test. Short speech samples (11–12 seconds) were presented to two groups of listeners over good quality loudspeakers in a lecture room. The task of the listeners was to rate the similarity between the mock perpetrator's voice and the voices of each the 15 possible foils using a 5-point scale. The listeners were also asked to estimate the age of all 16 speakers. Two groups of undergraduate students (37 altogether) took part in this test.

Inter-rater agreement was high for both age estimates (α = .964) and voice similarity (α = .933). There was no significant correlation between estimates of age or voice similarity and listener sex or age. *Chronological Age* and *Estimated Age* were significantly correlated ($p < .01$) but explained variance was moderate (R^2 = .155). *Estimated Voice Similarity* correlated significantly with *Estimated Age* ($p < .01$) but explained variance was low (R^2 = .038). There was no significant correlation between *Estimated Voice Similarity* and *Chronological Age*, however, indicating that voice similarity judgments to some degree may be influenced by perceived age similarity. Even though explained variance is marginal, the implication is that the age range of foils is a factor to be considered when constructing a line-up.

The number of foils suggested in other studies varies between five and eight with no general agreement on what the optimal number might be (Broeders & Amelsvoort, 1999; Bull & Clifford, 1984; Hollien, 1996; Hollien, 2002). For our study we decided to use six foils.

To ensure that no particular foil would stand out as "different" due to deviant perceived age we selected foils with a perceived age within a range of five years from the perceived age (25 years) of the perpetrator. Regarding perceived voice similarity, we selected two foils who had been rated as "quite similar", two rated "rather dissimilar" and the remaining two in the middle following suggestions by Hollien (2002). Combining these requirements gave us 6 speakers whose perceived ages ranged between 22 and 30 years of age. The similarity scores for the most similar voices were 3.1 and 3.5, the least similar ones scored 1.6 and 2.4 and the two in the middle 2.6, and 2.7 (5-point scale). Using the selected recordings, three sets of line-ups were constructed choosing speech samples from the 'walk through the town' recordings.

It makes some sense to expect a correlation between target voice exposure time and identification accuracy, and several studies have explored that possibility with stimulus content and duration ranging from vowel segments of 25 ms duration (Compton, 1963) to 8 min monologues (Orchard & Yarmey, 1995). There appears to be tendency for the longest durations to produce more hits in the target-present line-ups, while the target-absent condition seems relatively unaffected (Yarmey & Matthys, 1992). But the advantage of longer duration is partly counteracted by a higher degree of false alarms (Yarmey, 1991). For ecologically more realistic durations in the 30–60 s range two studies have found hit rates comparable to those found for much longer durations (Kerstholt, Jansen, Van Amelsvoort, & Broeders, 2004; Legge, Grosmann, & Pieper, 1984). Durations in this range have also been used in other studies comparable to the present one (Nolan & Grabe, 1996; Philippon et al., 2007). Based on considerations of ecological validity and the results from the above studies we decided to use a target stimulus in that the 30–60 s duration range. The manuscript we had prepared for the incriminating call produced recordings in that duration range and the selected target stimulus had a duration of 40 seconds. Two sets of line-up stimuli were prepared; a set of longer versions (M = 24.6 s, SD = 1.0 s) and a set of shorter versions (M = 12.6 s, SD = 0.9 s).

Construction of the line-ups

Three direct recorded line-ups were constructed for the voice recognition tests, with the target voice at three different positions in the line-up and the positions of the foils randomized. Three identical mobile phone recorded line-ups were also constructed.

Procedure

Overview

At the witnessing situation, the participants were exposed to an unfamiliar voice, either direct recorded or mobile phone recorded. Two weeks later (±1 day), each participant was presented with a voice line-up, consisting of direct recorded voices or voices recorded through a mobile phone, and then interviewed about the event.

Witnessing situation

To simulate a naturally occurring situation where you may overhear someone speaking without seeing the speaker, we instructed the participants to imagine themselves being in a clothes store, standing in front of the curtain to a dressing room waiting for their turn. To create a realistic feeling, a curtain had been hung from the ceiling and subjects were instructed to place themselves in front of the curtain. Loudspeakers presenting the recorded target were placed behind the curtain. Subjects were told that while waiting they would overhear something from behind the curtain, but were not told that it would be a voice. The presentation started with a mobile phone signal followed by a voice answering the call and talking to someone (not heard) for about 40 seconds about the planning of a crime. The content of the conversation was inspired by (but not identical to) a conversation used by Philippon (2006) and started with an utterance meant to catch the attention of the listener: "I have told you not to call me under any circumstances; the phone might be tapped!" (See appendix A). For half of the participants the voice was direct recorded and the other half heard an identical conversation recorded through a mobile phone. Thus, the only factor that differed between the two conditions was the sound quality of the recording. After the presentation the participants were told that they would be interviewed about the event two weeks later; however they were not informed about what aspects of the phone call the interview would concern.

The line-up

Two weeks after the event, the participants were confronted with a recorded voice line-up consisting of seven voices. Half of the witnesses were exposed to a direct recorded line-up, and the other half to a mobile phone recorded line-up. All line-ups were target-present line-ups. The participants were informed that the person they had heard speaking two weeks earlier may or may not be present in the line-up. Prior to each voice a paper with a number (1 to 7) was placed in front of the participants to inform them about the number assigned to each voice. First the participants were instructed to listen carefully to all seven voices represented by 22–26 second voice samples (M = 24.6). After hearing all voices once, the participants were instructed that they would now hear the voices once again, but shorter samples (11–14 seconds, M = 12.6) for each voice. This time the participants were asked to report if they recognised the voice they had heard earlier and if so, which of the numbered voices they thought it was. If they thought that the voice was not present they were instructed to simply say so.

The participants were also asked to report how certain they were that their decision was correct and how difficult they perceived the task to be. When asking the participants to assess their certainty, the interviewer showed a 5-point scale which consisted of numbers (1–5) and a written description for each number: *very unsure* (1), *not so sure* (2), *neither unsure nor sure* (3), *pretty sure* (4) and *very sure* (5). Finally, the participants were asked for personal background information such as age, place of birth and years of life in Sweden. The interviews, which lasted about 10 minutes, were recorded on a solid state audio recorder.

Results

The order of the foils was randomized between the different line-ups in order to minimize the possible influence of presentation order which might generate false identifications as a function of foil position. In our analysis of the results we have, however, not found any significant correlation between false identifications and position (Pearson r = –0.029, p = 0.749). Data for the different line-up orders have therefore been pooled in the subsequent analyses.

We also tested for possible listener sex differences. With respect to correct identifications there is a tendency for female listeners to perform somewhat better than

male listeners (14% vs. 9%) but the difference is not significant, $F = 0.757$, $p = 0.386$. The same tendency is found for false rejections (27% vs. 38%) but this difference is not significant either, $F = 2.043$, $p = 0.155$. Male and female listeners made almost the same proportion of false identifications (53% and 59% respectively). The difference is not significant, $F = 0.541$, $p = 0.463$.

In the statistical analysis we have considered that the subjects had 8 possible choices – 7 possible identifications (false or correct) plus "not present" (also false or correct). We have not found any convincing argument to limit the number of choices to 7.

Identification as a function of stimulus quality

The principle question of this study is the possible influence of the inferior sound quality of mobile phone mediated speech on identification accuracy. All three possible outcomes of the identification test – correct identification, false identification, and false rejection – contain potentially valuable information in this respect. The focus of this section will therefore be on the possible influence of stimulus sound quality.

Correct identifications

The overall mean for correct identification across conditions was 12.7%, which is almost exactly what would be expected by chance (12.5%, 8 possible alternatives).
As seen in Table 1, the number of correct identifications is somewhat lower for the direct recorded line-ups (in direct contradiction of Hypothesis 1), but none of the differences are significant. Stimulus sound quality does not seem to have had any influence on the results with respect to correct identifications.

Table 1. A summary showing the number of subjects in each condition (Total) and the number of responses for each possible alternative (ID).

	Presentation format					
	Direct			**Mobile phone**		
	Total	**ID**	**%**	**Total**	**ID**	**%**
Line-up: Direct						
Correct identification	42	4	10	43	4	9
False identification	42	24	57	43	27	63
False rejection	42	14	33	43	12	28
Line-up: Mobile phone						
Correct identification	39	6	16	41	7	17
False identification	39	20	51	41	23	56
False rejection	39	13	33	41	11	27

False identifications and false rejections

Across all conditions, 57% of the participants made false identifications. As shown in Table 1 there is little variation in the number of false identifications as a function of presentation format or line-up format and the differences are not statistically significant. As was the case for correct identifications we may thus conclude that stimulus sound quality does not seem to have influenced the results with respect to false identifications either.

One reason for selecting foils based on a voice similarity perception test was that we hoped to be able to use these data to explain any possible bias toward particular foils in the false alarm responses. It seems like a reasonable hypothesis that the foils that were perceived as sounding more similar to the target voice in the perception test would be mistaken for the perpetrator more often in the line-ups. In the statistical analysis of false identifications we did not find any significant correlation between the number of false identifications per foil and the scores obtained in the voice similarity test. We did, however, find a very strong bias towards one particular foil who was chosen significantly more often than any other foil and even the target speaker. 54% of the subjects who made false identifications selected this particular foil. The number of times this particular foil was chosen was significantly above chance level for all conditions ($p < .05$ for the direct presentation and mobile phone line-up and $p < .01$ for all other combinations). The results did, however, not correlate with presentation format

or line-up format. This foil scored right in the middle of the perceived similarity range so the explanation for this bias must be sought elsewhere. A possible explanation will be suggested in the discussion.

The proportion of false rejections was almost the same for all four conditions and the differences were not significant.

Confidence and perceived difficulty

The overall confidence-accuracy correlation was not significant, $r(165) = .002$, $p = .979$, which lends support for Hypothesis 3. Further a 2 (presentation: direct vs. mobile phone) x 2 (line-up: direct vs. mobile phone) ANOVA did not reveal a significant difference in reported confidence level across the four conditions, $F(3, 161) = 1.004$, $p = .392$.

Perceived difficulty was rated on a 5-point scale ranging from *very difficult* (1) to *very easy* (5). The mean perceived difficulty was 1.9 ($SD = .9$) and did not vary as a function of presentation format or line-up format, $F(3, 161) = 1.783$, $p = .152$.

There was, however, a significant correlation between confidence and perceived difficulty. Confidence decreased monotonously with increasing perceived difficulty, $r(165) = .524$, $p < .001$.

Discussion

The goal of the present study was to investigate if mobile phone quality has an effect on voice identification accuracy. Our study is, together with the study by Kerstholt et al. (2006), the first to address this question (previous research has used landline phones or telephone quality simulations).

Regardless of condition, the results showed that the number of correct identifications was at chance level. The hypothesis that the number of correct identifications would be higher in the direct conditions compared to the mobile phone conditions was not supported. In fact, the absolute number of correct identifications was slightly higher for the mobile phone line-up conditions but, there was no significant difference between any of the four combinations of presentation format and line-up format. One might ask if there could still be an influence of mobile phone quality but

that the effect was not detected due to the low number of correct identifications. Our analyses failed, however, to produce a significant effect of mobile phones not just for correct identifications, but for any of the obtained results including the significant number of false identification of one particular foil. Taken together these results strongly suggest that the detrimental effect on voice recognition suggested by the poorer sound quality of mobile phone recordings is minimal, if any at all. Similar results have been obtained in other telephone studies (Kerstholt et al., 2006; Perfect, Hunt, & Harris, 2002; Yarmey, 2003). Further our results also imply, in accordance with our Hypothesis 2, that using a mobile phone recorded voice line-up when the voice is originally heard over a mobile phone is not likely to improve identification accuracy. The absence of a detrimental effect of mobile phone quality on voice identification accuracy is somewhat surprising considering the clear sound quality differences. We have no conclusive explanation as to why the vastly inferior sound quality does not seem to have an effect on voice identification.

Regarding false identifications, a majority of the participants (57%) falsely identified a foil even though the target voice was always present. Witnesses' strong tendency to identify an innocent person is nothing new and also what previous research has often shown (e.g. Vanags, Carroll, & Perfect, 2005; Yarmey, 2003; Öhman, et al., unpublished). What was surprising, however, was that when checking for possible biases we found that of all participants who made false identifications, more than half of them (54%) chose one particular foil. This was quite unexpected since we had taken all the normal precautions meant to prevent such a bias from occurring.

Why this foil was so frequently chosen is an important question since such biases may severely affect recognition accuracy. In order to address this bias, and much inspired by the results obtained by Petrini and Tagliapietra (2008), we explored possible correlations with known acoustic correlates to speaking style and voice. We found that the most identified foil was the one that was most similar to the target with respect to articulation rate and pausing. If the subjects remembered the target as a fast talker then this perceived similarity is one possible explanation for the bias. However it is also possible that they chose that speaker only because the high rate attracted their attention in general. In a study by Zetterholm, Sarwar, and Allwood (2009), using a target-absent line-up, it was found that the foil that was most similar to the target voice with respect to articulation rate and pitch was the most often chosen foil. Such factors, which may severely distort a recognition task, should be addressed in future studies. Some recent

studies have also tested other voice characteristics that may have an effect on earwitness memory such as high-pitch vs. low-pitch voices (Mullennix et. al., 2009b) and high-typical vs. low-typical voices (Mullennix et al., 2009a).

In line with our prediction we found no significant confidence-accuracy correlation which is in accordance with previous findings showing that confidence is not a good predictor of voice identification accuracy (Kerstholt et al., 2006; Olsson et al, 1998; Yarmey, 1991, Yarmey, 2001; Öhman et al., unpublished). In contrast to Perfect, Hunt, and Harris, (2002) and Yarmey (2003), we did not find a significant difference in confidence judgements between those who heard a mobile phone recorded voice and a direct recorded voice.

The number of correct identifications was at chance level. This finding parallels findings in similar studies stressing the fact that people are not particularly reliable as earwitnesses. Exactly why this is so, is in need or further investigation, but some potentially relevant observations may be mentioned. The perpetrator's voice changed from hostile to non-hostile (normal tone) between presentation and line-up, and such a change in tone has previously been found to reduce identification to chance level (Read & Craik, 1995; Saslove & Yarmey 1980). Furthermore, witnesses in real-life situations are seldom prepared for what is coming and in order to approximate ecological validity our participants were not given any detailed information in advance about what they were about to experience. To be unprepared for the witnessing event has also previously been found to lower the results (Hollien, Huntley, Künzel, & Hollien, 1995; Saslove & Yarmey 1980; Yarmey, 2003). Due to the considerable variation in methodology between earwitness studies it is not possible to say exactly what level of identification accuracy one should expect for a given situation, but Read and Craik (1995) concluded that voice identification accuracy for an unfamiliar voice heard under realistic conditions is somewhere between 30% and down to chance level which is a level that is far from satisfactory in real life investigations.

Limitations

One could speculate that our low identification accuracy might have worked against a proper test of whether presentation format and/or line-up format moderate earwitnesses' performance. Furthermore, our analysis suggests that one particular foil may have contributed to this floor-effect (i.e., this foil was selected by 44% of all

participants making an identification, correct or false, as compared to 18% for the target). In brief, it might be the case that the inclusion of this particular foil in the line-up, to some extent, masked any 'true' effects of presentation and/or line-up format. It is for future research to show whether this speculation holds true.

The present study aimed to test if mobile phone sound quality caused by technical factors has an impact on voice identification accuracy, and did not include other factors connected with mobile phone speech like speaking style. For the current study, the direct and mobile speech samples were recorded simultaneously which most likely meant that the speakers did not have the feeling that they were speaking in a mobile phone and therefore did not adjust their speech the way they might normally have done. One factor that often influences mobile phone speech is ambient noise which in the normal case makes the speaker talk louder. It may well be the case that using mobile phones in a noisy environment is so common that mobile phone users automatically raise their voices also when this is not necessary to overcome ambient noise, but we know of no study which may confirm this assumption. If and how factors connected with mobile phone speaking style affect identification accuracy is an important question that should be tested in future experiments.

Conclusions, legal implications and future directions

The main finding in the present study is that mobile phone quality does not seem to have an effect on voice identification accuracy and that the idea that mobile phone line-ups should be used if the target voice is originally heard over a telephone receives no support.

Earwitness identification accuracy was poor. This finding is in line with previous studies showing that earwitnesses are not particularly reliable whether the voice is heard directly (Deffenbacher et al., 1989; Read & Craik, 1995; Saslove & Yarmey, 1980; Öhman et al, unpublished) or over a phone (Yarmey, 2003).

In our study, an unfortunate choice of one foil introduced a strong selection bias toward that particular foil. Whether the identification results would have been significantly better if a different foil had been used is impossible to say with certainty, but it seems likely that that might be the case. Could this problem have been avoided? As described above we took all the usual precautions following a selection procedure

almost all other studies have followed and in more or less the same way but this did not help. Certain aspects of speaking style seem to have been involved, like speaking rate, but this is unlikely to be the full explanation. It should also be stressed that these factors seem to have escaped the attention of the judges who rated the foils for similarity. Although quite a lot is known about various parameters which may be used to describe speaking style there are, however, at present no known procedures for selecting a balanced set of foils based on speaking style. An obvious conclusion is that these factors should be examined more closely in future studies.

Earwitnesses may play an important role in criminal investigations but for them to be really useful we must find ways of improving accuracy. Identification accuracy between chance level and 30% is simply not satisfactory from a legal perspective. Finding ways of improving accuracy should therefore be an important concern in future studies. However, it is important that the possible methods are realistic in real-life situations to ensure high forensic relevance.

Acknowledgements

This research was supported by a grant given to Anders Eriksson from the Crime Victim Compensation and Support Authority (Grant no. 03347/2007).

References

Broeders, A. P. A., & Amelsvoort, A. G. V. (1999). Lineup construction for forensic earwitness identification: A practical approach. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1373–1376). San Francisco.

Brümmer, N., & Strasheim, A. (2009). *AGNITIO's Speaker recognition system for EVALITA 2009*. Paper presented at the The 11th Conference of the Italian Association for Artificial Intelligence.

Bull, R. H., & Clifford, B. R. (1984). Earwitness voice recognition accuracy. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 92–123). Cambridge: Cambridge University Press.

Byrne, C., & Foulkes, P. (2004). The 'mobile phone effect' on vowel formants. *The International Journal of Speech, Language and the Law, 11*, 83–102.

Compton, A. J. (1963). Effects of filtering and vocal duration upon the identification of speakers aurally. *Journal of the Acoustical Society of America, 35*, 1748–1752.

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decisionmaking in eyewitness identification cases. *Law and Human Behavior, 12*, 41–55.

Deffenbacher, K. A., Cross, J. F., Handkins, R. E., Chance, J. E., Goldstein, A. G., Hammersley, R., et al. (1989). Relevance of voice identification research criteria for evaluating reliability of an identification. *Journal of Psychology, 123*, 109–119.

Foulkes, P., & Barron, A. (2000). Telephone speaker recognition amongst members of a close social network. *Forensic Linguistics, 7*, 180–198.

Guillemin, B. J., & Watson, C. (2008). Impact of the GSM mobile phone network on the speech signal: Some preliminary findings. *The International Journal of Speech, Language and the Law, 15*, 193–218.

Hirson, A., French, J. P., & Howard, D. M. (1995). Speech fundamental frequency over the telephone and face-to-face: Some implications for forensic phonetics. In J. W. Lewis (Ed.), *Studies in general and English phonetics: Essays in honour of Professor J. D. O'Connor* (pp. 230–240). London & New York: Routledge.

Hollien, H. (1996). Consideration of guidelines for earwitness lineups. *Forensic Linguistics, 3*, 14–23.

Hollien, H. (2002). *Forensic voice identification*. San Diego, CA: Academic Press.

Hollien, H., Huntley, R., Künzel, H., & Hollien, P. A. (1995). Criteria for earwitness lineups. *Forensic Linguistics, 2*, 143–153.

Kerstholt, J. H., Jansen, N. J. M., Van Amelsvoort, A. G., & Broeders, 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., & Broeders, A. P. A. (2006). Earwitnesses: Effects of accent, retention and telephone. *Applied Cognitive Psychology, 20*, 187–197.

Künzel, H. J. (2001). Beware of the 'telephone effect': The influence of telephone transmission on the measurement of formant frequencies. *Forensic Linguistics, 8*, 80–99.

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. (1944). An experimental study of voice recognition. *Journal of General Psychology, 31*, 53–65.

Mullennix, J. W., Ross, A., Smith, C., Kuykendall, K., Conard, J., & Barb, S. (2009a). Typicality effects on memory for voice: Implications for earwitness testimony. *Journal of Applied Cognitive Psychology*, Retrieved October 7 2009, from http://www3.interscience.wiley.com/cgi-bin/fulltext/122623678/PDFSTART

Mullennix, J. W., Stern, S. E., Grounds, B., Kalas, R., Flaherty, M., Kowalok, S., et al. (2009b). Earwitness memory: Distortions for voice pitch and speaking rate. *Journal of Applied Cognitive Psychology*. Retrieved March 2 2009, from http://www3.interscience.wiley.com/cgi-bin/fulltext/122212273/PDFSTART

Nolan, F., & Grabe, E. (1996). Preparing a voice line-up. *Forensic Linguistics, 3*, 74–94.

Öhman, L., Eriksson, A., & Granhag, P. A. (unpublished). Overhearing the planning of a crime: Do adults outperform children as earwitnesses? University of Gothenburg, Sweden.

Olsson, N., Juslin, P., & Winman, A. (1998). Realism of confidence in earwitness versus eyewitness identification. *Journal of Experimental Psychology: Applied, 4*, 101–118.

Orchard, T. L., & Yarmey, A. D. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology, 9*, 249–260.

Perfect, T. J., Hunt, L. J., & Harris, C. M. (2002). Verbal overshadowing in voice recognition. *Applied Cognitive Psychology, 16*, 973–980.

Petrini, K., & Tagliapietra, S. (2008). Cognitive maturation and the use of pitch and rate information in making similarity judgments of a single talker. *Journal of Speech, Language, and Hearing Research, 51*, 485–501.

Philippon, A. C. (2006). *Social, cognitive and methodological aspects in earwitness identification.* Portsmouth: University of Portsmouth,.

Philippon, A. C., Cherryman, J., Bull, R., & Vrij, A. (2007). Earwitness identification performance: The effect of language, target, deliberate strategies and indirect measures. *Applied Cognitive Psychology, 21*, 539–550.

Rathborn, H. A., Bull, R. H., & Clifford, B. R. (1981). Voice recognition over the telephone. *Journal of Police Science and Administration, 9*, 280–284.

Read, D., & Craik, F. I. M. (1995). Earwitness identification: Some influences on voice recognition. *Journal of Experimental Psychology: Applied, 1*, 6–18.

Rothman, H. B. (1977). A perceptual (aural) and spectrographic identification of talkers with similar sounding voices. In J. S. Jackson (Ed.), *International Conference on Crime Countermeasures - Science and Engineering* (pp. 37–42). Oxford: University of Oxford.

Saslove, H., & Yarmey, A. D. (1980). Long-term auditory memory: Speaker identification. *Journal of Applied Psychology, 65*, 111–116.

Vanags, T., Carroll, M., & Perfect, T. J. (2005). Verbal overshadowing: A sound theory in voice recognition? *Applied Cognitive Psychology, 19*, 1127–1144.

Van Wallendael, L. R., Surace, A., Hall-Parsons, D., & Brown, M. (1994). 'Earwitness' voice recognition: Factors affecting accuracy and impact on jurors. *Applied Cognitive Psychology, 8*, 661–677.

Yarmey, A. D. (1991). Voice identification over the telephone. *Journal of Applied Social Psychology, 21*, 1868–1876.

Yarmey, A. D. (1995). Earwitness speaker identification. *Psychology, Public Policy, and Law, 1*, 792–818.

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: Forensic Linguistics, 10*, 62–74.

Yarmey, A. D., & Matthys, E. (1992). Voice identification of an abductor. *Applied Cognitive Psychology, 6*(5), 367–377.

Zetterholm, E., Sarwar, F., & Allwood, C. M. (2009). Earwitnesses: The effect of voice differences in identification accuracy and the realism in confidence judgments. In P. Branderud & H. Traunmüller (Eds.), *Proceedings of FONETIK 2009, The XXIIth Swedish Phonetics Conference* (pp. 180–185). Stockholm: Stockholm University.

Instructions

Presentation

The *European Journal of Psychology Applied to Legal Context*, the Official Journal of the Sociedad Española de Psicología Jurídica y Forense, publishes empirical articles, theoretical studies and focused reviews of topics dealing with psychology and law (e.g., legal decision making, eyewitness). Only original papers (not published or submitted elsewhere) will be published. Papers driven to both legal systems, inquisitorial and adversarial, will be welcome as well as papers based in concrete laws of a European country. Neither the Editors nor Publishers accept responsibility for the views or statements expressed by the authors.

Paper submission

Manuscripts should be submitted electronically to the Editors to the e-mail address of the journal (ejpalc@usc.es). Postal address should be used exceptionally (The European Journal of Psychology Applied to Legal Context, Facultad de Psicología, Universidad de Santiago de Compostela, E-15782 Santiago, Spain). Submission of a paper to this journal implies that it represents original work not previously published, and that it is not being considered elsewhere for publication.

Review

The manuscripts will be reviewed by two external blind referees. The reviews are anonymous for authors and reviewers. Author identities will be removed before sending out a manuscript to the reviewers.

Copyright

Authors submitting a manuscript do so with the understanding that if it is accepted for publication the copyright of the manuscript, including the reproduction of the paper in all forms and media, shall be transferred to the publisher.

Permissions and responsibility

The author is responsible for obtaining permission necessary to quote from other works, to reproduce material already published, and to reprint from other publications. The opinions expressed and the contents of the paper are under exclusive responsibility of the author(s) and do not reflect the point of view of The European Journal of Psychology Applied to Legal Context.

Style

Manuscripts must be adhere to the instructions on references, tables, figures, abstract, format, narrative style, etc. as described in the Publication Manual of the American Psychological Association (5th edition). Manuscripts that do not fit to the style set forth in this manual will not be considered for publication.

Check list of requirements

The abstract should be 150-200 words.

Title page (include the authors´ name, affiliations, full contact details).

Full paper text (double spaced with numbered pages and anonymised).

References (APA style).

Tables and figures placed at the end of the paper or attached separately.

The European Journal of Psychology Applied to Legal Context

CONTENTS

Articles

Judicial judgement-making and legal criteria of testimonial credibility
Mercedes Novo and Dolores Seijo 91

Juvenile delinquency and young offender: bibliographical and bibliometric review of two perspectives of study
Francisco Javier Rodríguez, Luis Rodríguez-Franco, Javier López-Cepero, and Carolina Bringas 117

Modelling alcohol consumption during adolescence using zero inflated negative binomial and decision trees
Elena Gervilla, Berta Cajal, Joan Roca, and Alfonso Palmer 145

Mobile phone quality vs. Direct quality: How the presentation format affects earwitness identification accuracy
Lisa Öhman, Anders Eriksson, and Pär Anders Granhag 161

A stepwise approach to identify intellectual disabilities in the criminal justice system
Erik Søndenaa, Tom Palmstierna, and Valentina Cabral Iversen 183