

A PRACTICAL APPROACH TO FORENSIC EARWITNESS IDENTIFICATION: CONSTRUCTING A VOICE LINE-UP

A. P. A. BROEDERS¹, A. G. van AMELSVOORT²

¹ *Netherlands Forensic Institute, Rijswijk, the Netherlands*

² *Institute for Crime Investigation and Crime Science, Zutphen,
the Netherlands*

ABSTRACT: In the last decade there has been a growing interest in questions raised by the identification of persons by victims and witnesses based on their voices. In the absence of clear criteria for the design of earwitness identification procedures, a set of guidelines has been developed which is partly based on a tried and trusted method developed by the Dutch police and judiciary for the visual identification of persons by eyewitnesses. Parallels and differences between person identification in the visual and the auditory domains are illustrated. Some of the issues which needed to be resolved in designing the procedure laid down in the guidelines are: the choice between single and multiple formats, between repeated trials and single trials, the use of actors, the choice between verbatim text or free speech, and the choice between suspect similarity versus culprit description in the selection of foils. Critical elements of the procedure are described in detail. A condensed English version of the guidelines is available on request.

KEY WORDS: Voice line-up; Speaker identification; Identification by witnesses.

Problems of Forensic Sciences, vol. XLVII, 2001, 237–245

Received 3 January 2001; accepted 15 September 2001

INTRODUCTION

Forensic speaker identification is typically involved with what is essentially a verification task: it seeks to establish whether a person who is believed to be associated with a particular crime is the same as the person whose voice was recorded in connection with the commission of that crime. As also appears from some of the papers presented at the Forensic Speech and Audio Analysis Session of this Second European Academy of Forensic Science Meeting, various methods and techniques are currently utilised to address this type of question. These range from predominantly phonetic-linguistic approaches to semi-automatic pattern recognition-type techniques, with the most promising approach for the future probably being one which combines the strengths of these rather different methods [2, 4]. Obviously,

what all these approaches presuppose is the availability of a questioned recording of the perpetrator's voice as well as a reference recording of the speech of the suspect, preferably both made in similar technical and situational conditions.

In some cases, none of these approaches can be used for the simple reason that the perpetrator's voice was not recorded. What we may have instead is a witness or a victim who has heard the voice of the perpetrator. In this type of situation it may occur to the police that this witness may be able to identify the suspect of that crime as the perpetrator. Or, alternatively, if the witness fails to recognise the suspect's voice, this may suggest to the police that they are on the wrong track and may cause them to pursue a different line of investigation. Apart from its use as an investigative tool, an identification test may also be used for evidential purposes. However, such evidence should be treated with considerable caution. Witness identification in general is notoriously problematic, as appears again from a recent study of miscarriages of justice in the USA [7]. Wells et al. [16] list 40 cases of wrongly convicted persons who were exonerated by DNA evidence, of which no fewer than 36 (or 90%) involved eyewitness identification evidence in which one or more witnesses falsely identified the innocent suspect. Flawed identification procedures in particular are notorious for producing unreliable evidence.

PARALLEL WITH VISUAL IDENTIFICATION PROCEDURES

In some respects identification of speakers by earwitnesses resembles identification by eyewitnesses. The latter is not only by far the more frequently occurring type of person identification in the forensic context but also clearly the more widely studied of the two. In the last decades a considerable body of literature has been published on eyewitness identification procedures [see 16]. Many of these provide important insights and principles for the field of speaker identification by earwitnesses. To the extent that this seemed applicable and feasible, these principles and insights have been brought to bear on the guidelines we are proposing. The first foundation for these was laid in the "Handleiding confrontatie" [1], a manual for the administration of visual identification tests which has come to be used almost universally by the Dutch police since its first publication in 1994.

SINGLE VERSUS MULTIPLE CONFRONTATION

One of the most common procedures used in this context is the voice line-up or multiple auditory confrontation. It typically consists of a record-

ing of the speech of 5 or 6 similar sounding speakers, one of whom is the suspect of a crime, the other, similar sounding speakers being called distractors or foils. Of course, we could simply ask the witness to listen to a recording of the suspect's voice only, to see if recognition takes place. However, we know from laboratory studies in the visual domain that this is not a good idea. Although the number of hits, i.e. correct identifications, is comparable for the two modes, the disadvantage of the single format is that false positives cannot come to light, while they can in the line-up mode, except of course if the suspect is chosen. But if a foil is chosen, we normally know that this is an incorrect identification, at least if we have made sure that the foils all have iron-cast alibis. So, in general, the evidential value of an identification in a line-up will be greater than in a single confrontation because the line-up reduces uncertainty.

SIMULTANEOUS VS. SEQUENTIAL PRESENTATION

The choice between a single confrontation, involving the suspect only, as opposed to a multiple confrontation, consisting of the suspect plus a number of foils is a good example of the parallelism between the auditory and the visual fields. But there are also differences. One of these concerns the choice between simultaneous versus sequential presentation. Laboratory studies of visual line-ups suggest that performance is better, i.e. there are more hits and fewer false positives, if the line-up is presented sequentially [8]. Sequential presentation reduces the undesirable effect of relative comparisons, where witnesses may choose the most similar line-up member rather than make an absolute choice. Also, because false positives are known to be more frequent towards the end of the line-up, the sequential format is preferable because the witness does not know how many persons there are in the line-up. For the auditory line-up, where the choice between sequential or simultaneous presentation does not really arise, this leads to the recommendation not to tell the witness how many voices he or she is going to hear.

RELATIVE JUDGEMENTS SHOULD BE AVOIDED

One reason why relative comparisons are undesirable is that there are good reasons to believe that witnesses will tend to choose that person in the line-up who looks most like the perpetrator, regardless of whether or not the perpetrator occurs in the line-up [14]. Simultaneous presentation of a visual line-up, and repeated playback of a voice line-up will facilitate cross-reference and may thus encourage relative judgements.

UNDERLYING LOGIC

To appreciate the similarities as well as the differences between eyewitness and earwitness identification it is necessary to deal briefly with the underlying logic of the multiple identification procedure. Unfortunately, this underlying logic is not always fully understood, which may lead to less than optimal results in eyewitness identification. It is important to note that essentially what the eyewitness has is a mental trace of the appearance of the perpetrator. What we also have, or what we can elicit anyway, is a verbal description of the appearance of the perpetrator by the witness. Now, the idea is that the foils should all meet the verbal description given by the witness. However, the mental trace is likely to contain information over and above that contained in the verbal description. It is this information beyond verbal recall [14] that will enable the witness, and the witness only, to recognise the suspect as the perpetrator (assuming of course that the suspect indeed is the perpetrator).

SELECTION OF FOILS

Now what does this mean for the selection of foils? The first requirement is that the foil should meet the description given by the witness. At the same time though, the foils and the suspect should not be too similar: they should definitely not be clones, which would be carrying the suspect similarity principle to its (absurd) extreme. Clearly, this makes sense if we remember the underlying logic of the line-up. It is the information beyond verbal recall that will enable the witness to recognise the perpetrator, if he is there. In the words of Wells [14], one of the leading experts on eyewitness identification, it is this propitious heterogeneity or helpful diversity among the line-up members that is central to the underlying logic of the line-up. So, it follows that, in selecting foils it is conformity to the culprit description rather than resemblance to the suspect that should prevail as a criterion.

Now here is the rub! What happens if we try to apply this to the voice line-up? Descriptions of voices, or more correctly of speech samples, are notoriously problematic. Witnesses may not be able to provide sufficiently detailed and reliable descriptions of voices. They may not even be able to label the regional accent of the perpetrator. However, if we find that the suspect has a marked Amsterdam accent, we can hardly fail to use this as a criterion in selecting the foils. In general, we would expect voice descriptions by witnesses to provide an insufficient basis for foil selection. As a result, we will be forced to rely relatively heavily on suspect similarity as a criterion for foil selection.

But, as we have just seen, this potentially weakens the effectiveness of the identification procedure because it undermines the very basis of its underlying logic. Nevertheless, it appears that, for the earwitness line-up anyway, foils should not only meet the description of the voice by the witness. They should also meet the speech profile of the suspect as made by a competent phonetician, and be matched for features such as biological and social gender, perceived age, accent, dialect and the like. So, while the visual line-up can capitalise on the surplus of information in the mental trace of the culprit in the witness's memory, over and above that contained in the verbal description of the culprit's voice by the witness, the voice line-up necessarily has to settle for a high degree of suspect similarity. Inevitably, this tends to reduce the helpful diversity among the line-up members which makes the line-up an effective identification procedure.

OTHER CONSIDERATIONS

There is a host of additional considerations that need to be taken into account in constructing a proper line-up. Many of these can only briefly be touched upon in the following sections. A fuller discussion of some of these can be found in Broeders [3].

Fairness of the line-up

The notion of fairness is not always well understood in the context of line-up composition. A line-up is fair if it provides an optimal opportunity for the witness to recognise the perpetrator if he or she is there, while at the same time making it very difficult for any non-witness to pick out the suspect, regardless of whether the suspect is the perpetrator. If non-witnesses can be demonstrated to be likely to pick the suspect, selection of the suspect by the witness may be argued to be due to some other factor than recognition and may therefore not be assigned any added value. Note also that the notion of fairness does not require that the members in the line-up should all resemble the suspect as much as possible. This would increase the chance of a foil being picked instead of a guilty suspect.

Test listeners and mock witnesses

Evaluating the fairness of a line-up frequently involves the use of test observers, or test listeners in the case of a voice line-up, and mock witnesses. Test listeners may be asked to listen to a draft line-up prior to its administration, paying special attention to the presence of any speakers that stand out from the others. They should be aware of the nature of the crime, be fa-

miliar with the description of the perpetrator's speech and be from a similar linguistic and preferably socio-economic background as the witness. Knowledge of the nature of the crime may reveal the presence in the line-up of speech stereotypes associated with the crime in question. Clearly, we would not want any member(s) of the line-up to sound like a stereotypical rapist, extortionist or con artist to the witness, least of all if we were dealing with this particular type of crime. To assess the sociolinguistic balance of the line-up, we need a test listener who is an in-group member of the witness's subculture. Even so, accent and dialect assessments remain notoriously problematic. Speakers will quite frequently mark a line-up member's dialect or accent as deviating from what they see as the norm for their variety of the language. For this reason, professional dialectologists or linguistically informed native speakers of the language variety in question may need to be consulted.

If one or more members of the line-up are pointed out by the test observer as somehow conspicuous, they need to be replaced and the new line-up needs to be played to a new test listener until no more speakers are felt to stand out.

Mock witnesses are used in a variety of ways to assess the fairness of a line-up. They may be asked to identify the suspect, on the basis of the information contained in the description of the perpetrator's voice and the shared speech features which were used to select the foils, as well as information about the nature of the crime. Ideally, choices of line-up members are more or less equally divided over the line-up.

Nominal vs. effective/functional size

The nominal size of the line-up may not be a good indication of its actual size. In other words, it may be the case that some of the members of the line-up do not actually pose a realistic alternative to the suspect. If the perpetrator spoke French, a voice line-up containing a French speaking suspect and five Spanish speaking foils effectively amounts to a single rather than a multiple line-up. Various methods have been proposed to test whether the foils do in fact make a meaningful contribution to the line-up, involving notions such as effective line-up size and functional line-up size. Typically, line-ups are played to some fifteen to twenty mock witnesses. For a comprehensive discussion see Valentine & Heaton [13].

The use of actors and/or transcripts

Witnesses tend to indicate that recognition of the unknown speaker's voice would be easier, or possible only, if they could hear (some of) the same utterances as they remember the unknown speaker using. A study by Yarmey (in press) suggests that this does not actually improve recognition

performance in real-life like conditions. The so-called transcript method involves using excerpts from a recording of the suspect's voice made during a police interview which are later matched by identical utterances produced by actors. It appears that it is extremely difficult for actors to sound as natural and spontaneous as the one non-actor in the line-up [10, 12]. Quite apart from questions concerning the legitimacy of using a recording of the suspect's voice for this purpose, there is of course a very real danger that the suspect's speech during the interview may be very different from the type of speech the witness heard owing to differences in situational context. After all, speech is a behavioural biometric, which is subject to considerable variation with time.

Recommendations of the American Psychology/Law Society (AP/LS)

In response to repeated concerns expressed in the literature about flawed identification procedures a subcommittee of the AP/LS formulated a small number of what they consider to be essential guidelines for the organisation of visual line-ups. These guidelines were subsequently published in the form of an official Scientific Review Paper of the AP/LS [16]. The first of these guidelines stipulates that, in order to avoid even the very possibility of any form of bias in the presentation, the line-up should be administered as a double-blind test, with neither the person administering the test nor the witness having any knowledge as to where in the line-up (if at all) the suspect is situated. The second recommendation requires that the instruction should be unbiased in the sense that the witness is told the voice of the person associated with the crime may or may not be in the line-up. Foil selection is the subject of the third rule, which says that foil selection should be based on the verbal description of the perpetrator by the witness. The final recommendation relates to confidence statements by the witness, which if requested at all, should be made elicited and recorded before the witness knows whether he or she has identified the suspect. Confidence statements made in courtrooms, sometimes many months after the identification test took place and with hindsight are likely to be highly inflated [9, 15].

Two additional recommendations were not incorporated in the core rules. They are that the presentation should be sequential rather than simultaneous, and that the actual line-up and the eyewitness be videotaped. While the former is inevitable in the case of voice line-ups, the latter recommendation is clearly also one that is well worth implementing.

THE PROPOSED METHOD

So far, the number of voice line-ups carried out using the proposed method is limited. Stenberg [11] describes using it in a case in Sweden. A full presentation of the proposed method would obviously take too much space. However, a detailed description of the procedures currently used by the authors may be found in Broeders & van Amelsvoort [6] and is also available from the authors on request.

References:

1. Amelsvoort van A. G., Handleiding confrontatie, Elsevier, Hague 1999.
2. Boves L., Commercial applications of speaker verification: Overview and critical success factors, Proceedings of RLA2C Workshop on Speaker Recognition and its Commercial and Forensic Applications, Avignon 1998, pp. 150–159.
3. Broeders A. P. A., Earwitness identification: Common ground, disputed territory and uncharted areas, *Forensic Linguistics* 1996, vol. 3, pp. 3–13.
4. Broeders A. P. A., Forensic speech and audio analysis: the state of the art in 2000 AD, [in:] Actas del I Congreso de la Sociedad Española de Acústica Forense, Madrid 2000, pp. 13–24.
5. Broeders A. P. A., The role of automatic speaker recognition techniques in forensic investigations, Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm 1995, vol. 3, pp. 154–161.
6. Broeders A. P. A., Amelsvoort van A. G., Line-up Construction for Forensic Earwitness Identification: a Practical Approach, Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco 1999, pp. 1373–1376.
7. Connors E. [et al.], Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial, National Institute of Justice, Alexandria 1996.
8. Lindsay R. C. L., Wells G. L., Improving eyewitness identification form line-ups: Simultaneous versus sequential line-up presentations, *Journal of Applied Psychology* 1985, vol. 70, pp. 556–564.
9. Luus C. A. E., Wells G. L., The malleability of eyewitness confidence: Co-witness and perseverance effects, *Journal of Applied Psychology* 1994, vol. 79, pp. 714–724.
10. Nolan F., Grabe E., Preparing a voice line-up, *Forensic Linguistics* 1996, vol. 3, pp. 74–94.
11. Stenberg M., Desideria: a case report on identification by voice line-up, vide pp. 184–189.
12. Stuart-Laubstein A., Problems of voice line-ups, *Forensic Linguistics* 1997, vol. 4, pp. 262–280.
13. Valentine T., Heaton P., An evaluation of the fairness of Police line-ups and video identifications, *Applied Cognitive Psychology* 1999, vol. 13, pp. S59–S72.
14. Wells G. L., What do we know about eyewitness identification?, *American Psychologist* 1993, vol. 48, pp. 553–571.

15. Wells G. L., Bradfield A. L., “Good, you identified the suspect”: Feedback to eyewitnesses distorts their reports of the witnessing experience, *Journal of Applied Psychology* 1998, vol. 83, pp. 360–376.
16. Wells G. L. [et al.], Eyewitness identification procedures: Recommendations for line-ups, *Law and Human Behavior* 1998, vol. 22, pp. 603–647.
17. Yarmey A. D., Earwitness Descriptions and Speaker Identification, *Forensic Linguistics* 2001, vol. 8.