ON EARWITNESS LINEUPS

Harry Hollien PhD

Abstract

Some confusion, even controversy, has accompanied the concept and/or application of what has been referred to as voice parades -- or perhaps more properly: earwitness lineups. This concern appears justified because 1) they often are confused with eyewitness lineups, 2) not very much research has been conducted on how effective they are and 3) clear-cut procedures as to how they should be carried out -- and especially good standards for their conduct and interpretation -- are but minimally available. Yet they can be most useful to investigations/trials when a person is a victim, or an observer, of a crime wherein they did not or could not see the perpetrator but could hear him or her speak. This article will briefly summarize the status of earwitness identifications, review acceptable procedures, and set standards for their conduct.

Keywords: Voice parades; ear witnesses; ear witness identification; speakers; speaker identification; listeners; multiple choice trials; comprehensive review; foil talkers; distracter talkers; witness memory; voice identification

1 Institute for the Advanced Study of the Communication Processes; 63 Dauer Hall, University of Florida Gainesville, FL 32611; Email hollien@grove.ufl.edu; Tel: 352-378-9771
Introduction

Problems associated with earwitness lineups, or voice parades, are of some concern within the Law Enforcement and Judicial systems. Even though this form of speaker identification is becoming somewhat more common, it still is not always attempted, even when relevant. The reasons for this appear to result from three relationships. The first involves problems associated with eyewitness identification plus its influence on the structuring of voice parades. The second results from the research that has been carried out directly on earwitness lineups; it is both a little sparse and sometimes contradictory. Finally, the area suffers from the lack of robust structuring and adequate standards.

Defining Earwitness Lineups

Earwitness lineups or “voice parades” can occur when an individual – who was either a victim of, or witness to, a crime – heard the perpetrator but did not see him or her. The task then is for the victim to make judgments about the perpetrator’s identity by listening to voice samples obtained from a suspect (who may or may not actually be the criminal). To do so, a procedure must be employed which is both well-structured and sensitive enough to ensure that accurate judgments can be made -- and that it is fair to all parties.

An actual example wherein an earwitness lineup would be needed is where a woman was raped in a dark place by a man she could not see but, rather, heard. Later, when a suspect’s voice is brought before her for identification, all she has available to her is her auditory memory of the man she heard. It would not be a voice with which she was familiar, of course for, in that instance, she already would have identified him. Therefore, an earwitness lineup has to be created for her and, if she can identify him by his voice, he can be arrested and arraigned.

Another instance would be where an agent observing a suspicious site intercepted a phone call between two terrorists planning a bombing. Since this process appears to be quite similar to eyewitness lineups, the early structuring of voice parades rather closely paralleled them. Herein lies the first set of problems.

The Earwitness-Eyewitness Problem

It must be stressed that the resemblance of voice parades to eyewitness lineups actually creates nearly as many problems as it solves. First, there are issues with eyewitness procedures themselves, hence, that type of model can be misleading. Second, while accurate eyewitness identifications (live, photos, “mugbook”) can be quite important to the conduct of a criminal investigation or trial (Buckhout, 1974, Hollien, 2002, Loftus, 1979, Wells and Olson, 2006, Yarmey, 1995), they can be detrimental if misused or fallacious. Indeed, research projects addressing these issues have been conducted over many years. Even very early studies demonstrated that a number of problems exist. In one instance, a student jury “convicted a suspect” (accused of committing a simulated crime) only on the basis of testimony by a single eyewitness -- one who had only 20/400 vision and was not wearing glasses when the “crime was observed” (Loftus 1979). Even earlier, Buckhout and Figueroa (1974) demonstrated that they
could bias witnesses simply by misaligning certain photographs on a “spreadsheet” (mugbook) or just by the verbal instructions they provided. While these effects were not powerful, they were found to exist. Over the years, other studies have documented a variety of problems with eyewitness-based criminal convictions. Perhaps the most dramatic of these were where DNA evidence demonstrated that certain individuals -- convicted on the basis of eyewitness evidence - - actually were innocent (Shuster, 2007, US Department of Justice, 1999).

On the other hand, a substantial number of experiments have led to knowledge about conditions which support reasonably accurate eyewitness identifications. Figure 1 provides a brief listing of those factors which can upgrade (or, if negative, can downgrade) the process. Examples: investigators have found that 1) individuals generally are better at identifying members of their own race (Buckhout, 1974, Cross, Cross and Daly, 1971, Wells and Olson 2003), 2) females tend to be better at identifying females than males, whereas males do not show much of a gender bias (Cross et al 1971, Hollien 2002, Wells and Olson 2003), 3) individuals who are substantially more or less attractive than the general public appear easier to identify (Shepard and Ellis, 1973, Wells, 1993, Wells and Olson, 2003), as do people with faces that can be considered “distinctive” (Cross et al, 1971, Yarmey, 1993), 4) older individuals tend to be somewhat more identifiable than younger people (Cross et al, 1971, Hollien 2002) and 5) poor lighting, poor eyesight and other handicaps can reduce accuracy (Buckhout, 1974, Wells, 1993). Even children have been studied. Ross and his associates (2006) have provided information about how children’s’ groups – such as preschoolers and/or adolescents – can equal or even exceed adult performance for certain eyewitness tasks. However, as with other areas, these relationships vary with specific challenges.

As would be expected, accuracy rates tend to rise when conditions are favorable (see for example Cutler, Penrod and Martens 1987, Lindsay and Wells, 1985, Steblay, 1997, Wells, 1993, Wells, Memon, Penrod, 2006). In any event, information about this and related issues has been used to upgrade protocol for eyewitness lineups.

The second issue here involves the use of the eyewitness model as a basis for voice parades. In the past, anyway, a number of authors have argued in favor of this approach (Broeders and Rietveld, 1995, Clifford, 1983, Rietveld and Broeders, 1991; Yarmey, 1995). They posit that both involve using 1) sensory modalities, 2) memory, 3) confrontational situations, 4) the process of sorting a person from a group and so on. However, it also is clear that the two procedures are different with regard to how 1) auditory and visual memories are processed, 2) eyes and voices are structured, 3) fear, anger or arousal affect the two identification processes, 4) poor eyesight contrasts with hearing disabilities and 5) native abilities make certain people good at visual memory and others better listeners. These, and other differences, demonstrate that eyewitness and earwitness identifications do not simply mirror each other (Hollien, 2002, Hollien, Bennett and Gelfer, 1983, Hollien, Huntley, Kunzel and Hollien, 1995). Hence, it is important to focus on the fundamentals of the auditory speaker recognition process and how it can be structured.
Earwitness Basics

The primary problem here is that, in the past, many of the individuals who created voice parades did not also establish the specific processes and standards necessary for success (Hollien, 1990, Hollien and Majewski, 2009, Thompson, 1985). Thus, this presentation will address structure, establish overall standards and describe just how an earwitness procedure can be carried out. In doing so, it is recognized that a number of variables operate to affect the process. Moreover, they vary in severity, with some being virtually uncontrollable or at least difficult to control.

Difficult Conditions

Some elements associated with the voice parades are essentially uncontrollable. It can only be hoped that, when they do occur, their effects will be randomized, counter-balanced, or, at least, minimal. They include: 1) latency between confrontation and lineup (see below), 2) relevant training (see below), 3) memory for human voices (see below), 4) talent levels (DeJong, 1998, Hollien, Majewski and Doherty, 1982, Kuenzel, 1995), 5) chronological age (individuals are more accurate in identifying people in their own age range and, overall, the precision exhibited by children and the elderly can vary substantially with these tasks), see: Huntley, Hollien and Shipp, 1987, Huntley and Pass, 1995 and 6) when foreign languages or dialects are present.

Latency. Significant degradation in auditory identification accuracy can be expected to occur within but a few weeks, even though some auditors are capable of performing well after even longer periods of time (Clifford, Rathborn and Bull, 1981, Kersholt, Jansen, VanAmelsoort and Broeders, 2004, Kunzel, 1994, Papcum, Kreimon and Davis, 1989). Unfortunately, however, memory for voice eventually will decay and will do so even for the most talented of witnesses. In any event, any attempt to reduce the latency time between the incident and a voice parade is desirable.

Memory/Training. There are some factors/conditions which would suggest that many witnesses have inherent capabilities which permit them to effectively handle the earwitness task. Data here is provided by DeJong (1998) who studied the effects of memory, auditory capability and musical skills on earwitness accuracy. She asked if any of these factors could affect a person’s ability to identify witnesses from voice -- and if so, how. She reported that intelligence (as measured by cognitive processing) was a better predictor of a subject’s ability to identify speakers than were basic auditory and memory skills. She also found that listeners who exhibited a high degree of musical aptitude (and training) also performed better than others. Her findings have been supported (and greatly expanded) by Kraus (1995, 2009) and her associates who report that most humans exhibit unexpected sophistication in auditory processing and that training in music and/or languages -- plus focus – greatly enhances these capabilities (Bronkhorst,
2000, Parbery-Clark, Skoe, and Kraus, 2009, Strait, Skoe, Kraus and Ashley, 2009). In addition, researchers at the University of Florida have found that many untrained individuals are capable of making surprisingly accurate identifications even under severely challenging conditions (Hollien et al, 1982), and that they tend to be more accurate when stressed or aroused (Aarts, 1984). Without question, these factors can operate to increase the accuracy of identification.

The discussion above pretty much outlines the more difficult challenges facing an earwitness lineup administrator (see also Home Office, UK 2003). It would now appear useful to describe just how a voice parade can be structured. Its basic elements will be addressed as will standards. To do so, however, it first would appear necessary to describe the two major processes which are currently being employed.

**Earwitness Lineup Structure**

The two leading approaches to earwitness lineups can be identified as the traditional, multiple-choice or “simultaneous” lineup, and the continual, comprehensive review or “sequential” approach. Multiple choice or simultaneous lineups are where the suspect’s voice is added to sets of foil samples -- and groups (20-25) of these “sets” are presented to the witness. Comprehensive review or sequential presentations are where the witness listens to the samples individually and at his or her own pace; then is allowed to go back and review any of them that he or she desires to check before making a decision. Unfortunately, the two terms of simultaneous and sequential also have been adopted by the eyewitness practitioners -- and they are using them in a somewhat different manner. Hence, to avoid confusion, the alternate – and more descriptive -- terms of Multiple Choice and Comprehensive Review have been adopted by the present author.

**The Multiple Choice Lineup**

This approach is structured by having one of several relatively short speech samples (exemplars) produced by the suspect, embedded somewhere within a sequence of samples spoken by distracters (or foils). A recording is made of 20-25 sets of these samples but where the suspect and the foils are placed in differing orders. Speech sample content can vary over the trials also but it must be consistent for all speakers within a trial. The entire recording (i.e., sets of S plus F’s) is then played to the witness and he or she attempts to pick the perpetrator’s voice out of each set as it is presented. Once the entire trial is complete, the hits and misses are counted and a judgment made -- that is, if it is possible to do so. Various forms of this approach have provided the basis for much of the research in this area; it is still commonly used. If a score of 85% or better is obtained, an identification probably can be made.

It must be said, however, that this approach has been challenged by Broeders (1996) and others. They argue that these repeated presentations might result in the witness becoming familiar with the suspect’s voice -- and, hence, incorrectly selecting it as the perpetrators. Of
course, if the suspect is not also the criminal this would be unfair to him or her. However, since all the foils also are present over the total set of trials, the witness might rather “lock on” to one of them -- also in error, or not do so at all. Second, the cited group argues that only a single set of voices should be presented -- as in an eyewitness lineup -- since, in the latter case you only get to see the suspect and the foils but once.

The first argument suffers from the fact that no research has specifically shown that this “locking in” on a voice actually occurs -- that is, if the presentation is properly structured (see: Wretling, Sullivan, Schlichting, 1999). Neither does the second argument appear justified as many examinations (i.e., “step forward”, “turn left”, “turn right”, etc.) actually are carried out during eyewitness lineups. Hence, it cannot be said that the suspect is presented only once. In any event, even though these arguments do not appear to be very convincing, this approach still is a little awkward to carry out.

The Comprehensive Review Procedure

This approach is clearly the more powerful of the two. It appears to have been first proposed by Mayor and Komulainen (1989) but, in various forms, it has more-or-less provided the basis for a number of subsequent approaches (Nolan and Grabe, 1996, Nolan, 2003). Moreover, it has been further developed and used by the present author (Hollien, 2002, Hollien, Huntley, Kunzel and Hollien, 1995). In its present form, it can be best understood by consideration of Figure 2. Note that the witness room is relatively small and contains but two desks, two chairs and a mounted video camera which can provide a view of both the witness and the operator (of course, a regular two-room “lineup” suite with a one-way mirror also can be used if the audio and lighting are good). In any event, the witness sits at a desk with a recorder. The person (operator) at the other table provides a series of recordings; they are numbered serially and are of the suspect’s voice plus several foils. Each is on a separate recording. Note that the operator is not permitted to know which of the several recordings contains the suspect’s voice exemplar. The procedure to be followed is that the witness calls out a number and the operator gives him or her that number recording. It is then played (the witness is allowed to make notes if he or she wishes; the operator is not permitted to do so, or to see those made by the witness, or to comment). The rest of the recordings are then played in turn -- and in any numerical order desired by the witness. Once all the recordings have been heard, the witness is permitted to request (and replay) any of them he or she wishes and as many times as desired. Ultimately, the witness is asked to identify the suspect if possible and, if so, to specify which recording contains the “remembered” voice. To reiterate, the witness is not required to make a selection if unable to do so. Finally, please note again the video camera seen in the figure. It is by this means that relevant individuals can observe the witness’s behavior without being invasive. Further, a video recording permits the entire session to be permanently captured for possible future use. This approach also permits the observers to track the witness’s behavior without introducing any of the possible negative effects (pressure, biasing, intimidation, etc.) that could result from a more invasive procedure. Incidentally, the two rooms may be adjacent (with or
without a one-way mirror) or they may be remote from one another with only the video linking them. Finally, the use of earphones (by the witness) is recommended.

**Procedural Standards**

To reiterate, a voice parade is where a witness is required to identify (if possible) a remembered voice from a field of voices. Each must involve only a single witness and a single suspect (who, of course, may or may not be the perpetrator). If there is more than one witness, or more than one suspect, multiple lineups **must** be carried out – and *independently* of each other. Finally, the witness must have heard the suspect’s voice (for example, heard his utterances during a telephone call or heard him speak while being blindfolded) but not have seen him. In the latter case, both ear- and eye-witness lineups should be held. Naturally, none of the above are necessary if the witness already knows the perpetrator or his voice.

**Operational Standards**

a) **Site.** The site of the lineup has been described.

b) **Instructions.** Clear instructions to all parties also have been indicated.

c) **Parity.** Earwitness lineups should be scrupulously fair to both the witness(es) and the suspect(s). Hence, operational guidelines (as seen below) should be rigorously followed.

d) **Records.** All aspects of the process should be properly recorded. Included should be records of 1) the witness background and statements, 2) the source of the suspects and their characteristics and 3) all activities/phases related to the identification process. The records specific to the lineup should be independent of, and in addition to, routine records.

e) **The Witness.** There are a number of issues involving the witness and they must be addressed before the voice parade can be carried out. First, the witness should be required to demonstrate the competency to carry out simple speaker identification tasks. That is, he or she should be interviewed in an effort to determine if they attended to the perpetrator’s voice well enough to remember it (did the perpetrator talk a lot, shout, threaten, speak with an accent, etc.?). The witness also should exhibit hearing adequate for the listening task. Pretests may have to be administered in order to establish these competencies. Second, it often has been observed that a witness will assume that the criminal’s voice will be among those in the lineup (Bull and Clifford 1999, Hollien, Huntley, Kuenzel and Hollien, 1995, Yarmey, 2007). That is, they usually assume the perpetrator is among the “suspects”, if for no other reason than they were asked to participate in a lineup. This bias must be countered by assuring the
witness that the voice lineup may or may not include the alleged criminal. Third, one of the problems encountered in earwitness lineups involves assessment and control of the witness’s psychological state. Even though it is well known that psychological stress (and other emotions) can affect a person’s behaviors (Hicks and Hollien, 1981, Scherer, 1986), the exact nature of these effects has not been established. Indeed, no control is possible with respect to the emotional states during the confrontation. There is, however, some evidence that high emotional states lead to somewhat better auditory identification (Aarts, 1984). On the other hand, the examiner should interview the victim about this event, as learning something about what the witness originally felt can assist in enhancing lineup structure and permit the better interpretation of responses. Finally, it is also important to identify the psychological states or emotions the witness is experiencing during the voice parade. Just as with the above discussion, attempts should be made to identify and mitigate them if at all possible.

**Test Structure**

The precision with which an earwitness lineup is structured is critical. Especially important is how the samples are constructed and how the foils are selected.

**Stimuli.** First, the exemplar samples (i.e., the utterances by the suspect and the foils) should be equal in length and of the same fidelity. That is, it has been observed that identification accuracy is superior with longer samples (Hollien, 1990, Huntley and Pass, 1995, Orchard and Yarmey, 1995, Roebuck and Welding, 1993, Yarmey, 1995). A sample of 1-2 minutes in length, which includes several different types of utterances (i.e., dialog involving conversation, repeated phrases and oral reading, as well as extemporaneous, stressed and shouted speech), is usually adequate. They also should be long enough to provide witnesses with a reasonable repertoire of the suspect’s (and the foils’) speech. All samples should consist of approximately the same material. Second, reasonable control over the speech can be exercised here even if it was impossible to do so during the original confrontation between the witness and the suspect. Good recording fidelity should be employed equally – across all “lineup” samples. However, if the perpetrator actually was heard over a communication link with limitations, an attempt should be made to mirror its channel characteristics (Hollien et al, 1995, Kunzel, 1994, Nolan, McDougall and Hudson, 2008, Yarmey, 1994, 2003). Data from Nolan et al’s 2008 study indicated that high quality samples matched to similar quality exemplars provided the best accuracy; telephone-telephone was next best, with mixed channel combinations the poorest. Third, it is desirable to include at least two types of (speech) materials. They are: 1) neutral and text independent speech (for example, extemporaneous speech in response to unemotional but structured questions) and 2) text dependant words, phrases, and sentences. Repetition of the phrases reported by the witness as having been heard in the original confrontation also can be useful. It is also necessary to
carry out reasonably long exemplar sessions (Hollien, 1992) as any attempt by the suspect to disguise his or her voice can be mitigated by this procedure. The samples also should include instances of loud and/or stressed speech (see Blanchard and Foulkes, 2006). Shortcuts should not be taken. A case in point results from a study conducted by an investigator (Laubstein, 1997) who adapted actual interrogations of the “suspects” (by detectives) for the earwitness samples and then had those texts also read by the foils. Two studies were carried out, one using actors (who also heard the “suspect’s” speech) and the second using policemen as foils. In all cases, listeners were able to identify the presumed criminal simply by being asked to do so. The author appears to feel that her data argues against administering earwitness lineups at all. Actually, the information simply provides additional support for the standards established in this article.

Foils. No matter what approach is employed, the target voice must be presented within a field of other voices. Moreover, both the number and nature of these foil or distracter voices is important. Ordinarily, between five and seven foil speakers should be employed. They should be of approximately the same age as the suspect and generally exhibit the same dialect/accent. It is also desirable that they have roughly similar social, economic, and educational backgrounds. Actors, and individuals who speak quite differently from the suspect, should be avoided. The characteristic of the suspect’s voice may be described to the foils but they cannot be allowed to hear it. In short, the witness should be provided with a reasonable repertoire of (different) distracter voices. Further, while it is not sensible (or fair) to select all voices which sound markedly unlike the suspect, it also is unwise to use only individuals with voices which sound similar, or quite similar, to that person (Broeders, 1996, Hollien, 1992, Hollien, 2002). Indeed, sound-alikes (e.g., brothers, fathers, sons) should be avoided. To be impartial with respect to both the witness and the suspect, the lineup should contain a variety of voices with one perhaps somewhat similar to the target voice, yet another which is quite different and the rest more or less “in between” (Broeders, 1996, Hollien et al, 1983).

Finally, a couple of items should be stressed. First, all samples presented to the witness should be uniform with respect to the speech sample content. Second, it must be recognized that the perpetrator may have exhibited stress, excitement, loud speech or other emotions during the commission of the crime. If such behavior is reported by the witness, the person taking the suspect’s exemplar should require him or her to include this type of speech in the sample. Finally, there is little question but that attempts by a suspect to disguise his or her voice (even accidentally) can reduce witness accuracy (Hollien, 1990, 2002, McGlone, Hollien and Hollien, 1977, Huntley and Pass, 1995, Reich 1981). Even imitating voices can create problems of this type (Schlichting and Sullivan, 1997). These efforts may be controlled to a great extent when the exemplar is taken even though nothing can be done about any “voice disguise” which might have taken place during the original confrontation. Important
here is that the person who conducts the exemplar is knowledgeable about both speech and voice disguise, and how to counteract them (Hollien, 1992).

**A procedural test.** The final step to be taken before the voice parade is administered is to evaluate it. That is, one or more mock trials should be carried out once the recordings are complete. They should be presented to 4-6 dispassionate listeners who are told to either select the person who is “different from the others” or “who sounds like a criminal.” If these judges consistently identify either the suspect or one of the foils, the recordings cannot be considered unbiased and should be restructured. These auditors also should report (for correction) any noticed characteristic or flaws that might detract from the test’s integrity.

**Test Administration**

The standards for test administration have been discussed in detail above. Hence, they are now listed only in summary form:

1. The Comprehensive Review approach is preferred.
2. A clear and complete set of instructions must be provided.
3. Only high quality recordings should be used. However, protocol should take into account how the suspect was heard (free field, telephone, etc.)
4. Samples and their presentation should be carefully controlled. They all should be structured/presented in the same manner.
5. While a witness should be briefed about the procedure, no information about his or her performance should be provided until all aspects of the lineup are completed.
6. No partiality toward either the witness or suspect should occur. The entire lineup should be observed (remotely) by individuals representing the witness, suspect and agency.
7. The witness will be asked – but not required – to make an identification.

**Discussion**

These guidelines and standards, while not all encompassing, nonetheless include most of the necessary criteria for the development of acceptable eyewitness lineups. A model such as this can be used to signal personnel working in the field that they must start taking cognizance of good standards and robust procedures, as well as respond positively to new criteria, procedures
and/or research findings. That is to say, there is still room for upgrading the protocol described above. For example, research should be carried out in order to seek information about: 1) which of the classes of stimulus materials are most suitable (for example, should the suspect’s speech be text dependant and/or independent; neutral and/or stressed; extemporaneous and/or read), and 2) how practitioners should be trained – plus, how much training should be required. In short, earwitness lineups are a reality; they occur all over the world. As a consequence, efforts must be made to intelligently deal with them and for those individuals who wish to conduct them to meet acceptable standards.

References


Cross, J.F., Cross, J. and Daly, J. (1971) Sex, Race, Age and Beauty as Factors in Recognition of Faces, Perception and Psychophysics, 10:393-396.


**Figure Legend**

**Figure 1-** Some of the features that, when favorable, tend to increase the likelihood of accurate eyewitness identifications. A similar, but somewhat less complete, set could be provided for earwitness lineups.

**Figure 2-** Drawing of the setup for a “Comprehensive Review” type of earwitness lineup. The facility could consist of adjacent offices or interrogation rooms equipped with a video link or a one way mirror (with audio) or a lineup room with similar links. 
A portrays the witness, B = the recorder, C = the set of recordings containing the speech samples, D = the operator and E is the video camera. Found in the other room would be F, which are the TV monitors and G, the observers. This configuration is recommended for voice parades.
## Figure 1

<table>
<thead>
<tr>
<th>Suspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Attractiveness</td>
</tr>
<tr>
<td>Unusual Features</td>
</tr>
<tr>
<td>Age</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clear Viewing Conditions</td>
</tr>
<tr>
<td>Good Lighting</td>
</tr>
<tr>
<td>Distance</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Witness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control of Emotions</td>
</tr>
<tr>
<td>Good Eyesight</td>
</tr>
<tr>
<td>Good Memory</td>
</tr>
<tr>
<td>Short Delay to Lineup</td>
</tr>
</tbody>
</table>