

# BUILDING VOICE LINEUPS: ONE METHOD'S BIAS PROBLEMS

Ann Stuart Laubstein

Department Linguistics, Carleton University  
Ottawa, K1S 5B6

## 1. Introduction

It is readily apparent that accurate voice recognition is a common phenomenon and misidentification, although less common, also takes place. How this occurs - which features of the voice are instrumental in this recognition or misidentification - remains an empirical question.

Nonetheless, despite this lack of knowledge, voice identification is being used as evidence in suspect identification by the legal systems of both the United States and Canada. There is no well established method for getting witnesses to identify suspects' voices but the most common method involves the presentation of a voice lineup. This poses further problems. The major problem posed by the so-called voice or earwitness lineup is the fact that a voice lineup is not simply a voice lineup but a speech lineup. And there is much more content to speech than the simple sound of the voice. The critical criterion for a good lineup is fairness. There must be nothing that distinguishes the suspect from the foils - something that is remarkably difficult to ensure in a voice lineup.

This paper describes an experimental study which investigates the feasibility of using what shall be termed the "transcript method" in the preparation of voice lineups - the currently accepted procedure in Canada and the United States (Mayor, 1989). This method has been used in an attempt to eliminate the speech content from the voice lineup. It was for instance, the method chosen by the Ottawa Police in a recent investigation. There are a number of problems with this method - in particular, it is not clear that it is capable of producing an unbiased lineup.

## 2.1 The Method

The procedure for preparing voice lineups used by the Ottawa Police is typical and will serve as the basis for the description which follows. First the suspect is interviewed by the police and this interview taped. Subsequently the interviewer's questions are spliced out, leaving only the suspect's speech on the tape. A written transcript is then made of the resulting monologue. Actors are chosen and they are asked to read the transcript; in the normal case they do not hear the tape of the suspect, however in the Ottawa case they did. The transcript and, in this case the tape, were used to prepare the foil samples.

Because the suspect's tone was extremely aggressive and he spoke very fast it was felt that the actors should be afforded the opportunity of hearing the tape once, just to "get a feel" for the suspect's manner of speech. The actors were then asked to read the transcript aloud a number of times and their "best effort" was chosen.

## 2.2 Reasons for Using the Transcript Method:

The transcript method is used in an attempt to eliminate the variability found in a speech sample which is due to factors other than actual voice characteristics. Besides the more obvious information regarding gender and age, speech may point to diverse factors such as educational level, socio-economic background, ethnic heritage, occupation, health, and regional background. In addition, speech has different kinds of informational content - so-called "sentence meaning" and "speaker meaning", as well as affective content.

Controlling for all of these features, all of which may potentially threaten the lack of bias in a lineup, is fraught with difficulties. The solution was the transcript method.

## 2.3 Problems with the Transcript Method

As pointed out above relatively little is known about the identification of people on the basis of their voices. A set of characteristics has been identified which will lead to quite reliable machine identification (Hollien, 1990) but it remains an empirical question as to how this set relates to the parameters chosen by the brain.

Secondly there are all the problems associated with the physical characteristics of the machines used and the varying acoustical properties of the rooms in which the taping takes place. In addition, splicing out the interviewer's voice can have quite enormous effects on the recording.

These problems can be dealt with but it is not clear that the biggest problem - the inherent difference between the reading and speaking modalities is capable of solution.

The experiments described below investigated two voice lineups prepared by the Ottawa Police in a recent criminal investigation. The investigation was aimed at determining whether the lineups were unbiased and hence whether their use by the police in suspect identification would be legitimate.

### 3.1 Experiment 1: Materials

A voice lineup was prepared using the transcript method detailed above. In addition to the suspect's voice there were seven foils' voices - actors from an Ottawa theatre company. The suspect's "monologue" lasted just under a minute and the actors' versions were all within five seconds in length of the suspect's tape. The versions judged "most natural" and the suspect's voice were put together on a single tape in quasi-random order with a two minute pause following each voice.

A questionnaire of sixteen questions was prepared relating to a number of sound/speech characteristics and possible judgements based on voice quality such as loudness, pitch, propensity to violence, suspect versus foil judgement).

### 3.2 Procedure

Subjects were told that they would hear a tape of eight voices one of which was the voice of a suspect in an ongoing police investigation. They were told that their participation would be of use to the justice system in that they would be helping to judge the fairness of a voice lineup. They were asked to evaluate each of the voices on a scale of 1 - 5 on each of the 16 questions asked. They were told that they would have two minutes following each voice to answer the questions but if they wished they could respond to any questions during the time that they were actually listening to a voice. Each subject was given a small booklet with the instructions on the front page and each of the subsequent pages labelled SPEAKER 1 (2,3,...8). The two runs of the experiment took place in two different classrooms at the university. Subjects listened to the tape on a ghetto blaster. Each run of the experiment lasted less than 25 minutes.

### 3.3 Subjects

Subjects (n=72) participated voluntarily for bonus points towards their final exam marks. They were from two introductory linguistics classes, n= 41 (Group A), and n=31 (Group B).

### 3.4 Results

Means were calculated for each voice with respect to each of the 16 questions. T-tests were then run comparing the suspect's mean to the foil mean closest to him on each of the 16 parameters. The suspect differed significantly ( $\alpha \leq .05$  (Bonferoni  $\leq .003$ )) from all the foils in both groups on only one of the questions, q8. In Group B the suspect also differed significantly on questions 1 and 16. In group A these two questions approach significance ( $p = .028$  and  $.018$ ).

### 3.5 Conclusion

Figure 1 indicates the two groups are virtually indistinguishable from one another and questions 1 and 16 have virtually the same means. The lineup appears biased regarding #1, #8 and #16.

q1 How difficult was it to understand the speaker?

Easy 1 2 3 4 5 Difficult

q8 Did you notice any sounds on the tape besides the speaker's voice?

None 1 2 3 4 5 Many

q16 How sure are you that this speaker is the real suspect and not a foil?

Foil 1 2 3 4 5 Suspect

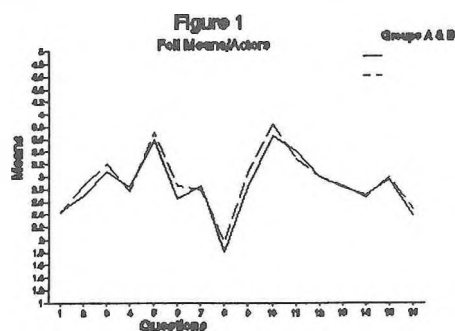
To overcome the biased nature of this lineup a new voice lineup was put together and a second experiment was run to determine whether it was biased.

### 4.1 Experiment 2: Materials, Subjects, Procedure

The tape was prepared in the same way as the tape for the first lineup except that the foils were Ottawa policemen and there were only six, not seven. The suspect was the 5th voice on the tape. The door and chair noises were included in all the foil samples. Subjects were again from two introductory linguistics classes (Group C n= 43, Group D n=47,) participating for bonus marks. The procedure was as in Experiment 1.

### 4.2 Results

The suspect's mean was significantly different from the closest foil on only two questions, #1 and #16 in both groups.



### 4.3 Conclusion

The lineup would again seem to be biased. It may be that this method is incapable of producing an unbiased lineup.

### References:

1. Hollien, H. (1990) *The Acoustics of Crime The New Science of Forensic Phonetics*, Plenum Press, New York.
2. Mayor, D. and Eeva Komulainen (1989) *Subjective Voice Identification ms. Calgary Police Service, Calgary, Alberta.*