BEHAVIOURAL SPEAKER IDENTIFICATION A FORENSIC APPLICATION

(Application de l'identification d'un interlocuteur par son comportement)

Steven J. Aiken, Donald G. Jamieson, and Vijay Parsa

Hearing Health Care Research Unit, The University of Western Ontario
School of Communication Sciences and Disorders, Elborn College
London, Ontario, CANADA, N6G 1H1

ABSTRACT

Behavioural speaker identification refers to the process of identifying an individual as the speaker of a given utterance, based solely on auditory perception. The present study applied behavioural speaker identification to evaluate the hypothesis that a particular individual had produced a set of utterances recorded in a police telephone tap. The individual admitted producing some of these utterances but denied producing others. Forty voice samples were extracted from 16 telephone calls recorded by the police, and one call with the individual, recorded by the experimenters. Two listeners rated the similarity of these samples in a paired-comparison task. Utterance pairs were grouped into pairs of potential speakers. An ANOVA was performed on the paired-comparison ratings with potential speaker pair and rater as factors. The effect of potential speaker pair was significant, suggesting that at least some of the utterances had been produced by different speakers. Utterance pairs in which both utterances were denied by the individual were rated as most similar, while utterance pairs in which only one utterance was denied by the individual were rated as least similar. A cluster analysis revealed two distinct clusters. All utterances denied by the individual fell into one cluster, while the other cluster was comprised of all utterances which the individual admitted producing, along with the recording of the individual's voice obtained by the experimenters. Overall, the evidence suggested that the individual was accurate in identifying which utterances he had produced.

SOMMAIRE

Identifier un locuteur par son comportement réfère au processus d'identification d'un individu qui dit un énoncé et ce, basé seulement sur l'information auditive. Dans la présente étude, l'identification d'un locuteur par son comportement a été appliquée afin de déterminer si un individu en particulier avait prononcé un énoncé potentiellement incriminant sur des enregistrements effectués par la police. Le suspect en question a admis avoir prononcé certains énoncés, mais nié en avoir produit d'autres. Quarante échantillons de voix ont été extraits de seize conversations téléphoniques enregistrées par la police et un appel au suspect a été enregistré par le chercheur. Des auditeurs ont noté les similarités de ces échantillons en effectuant des comparaisons par paires. Les paires d'énoncés ont ainsi été groupées en paires de locuteurs potentiels, en supposant que les énoncés d'un seula ppel ont été produits par un seul interlocuteur potentiel. Un ANOVA a ensuite été exécuté sur la comparaison des paires notées selon les paires de locuteurs potentiels, avec les évaluateurs comme facteurs. L'effet du locuteur potentiel s'est avéré significatif, ce qui suggère que certains des énoncés ont été produits par des locuteurs différents. Les paires d'énoncés comprenant deux énoncés niés par le suspect ont été notés de façon plus similaire que les paires où seulement un des deux énoncés avait été nié par ce dernier. Ainsi, une analyse de groupe a révélé deux groupes distincts. Tous les énoncés niés par le suspect se sont retrouvés dans un groupe, tandis que l'autre groupe se composait de tous les énoncés que le suspect avait admis avoir produits et ce , selon l'enregistrement du suspect obtenu par le chercheur. En somme, les résultats suggèrent que le suspect avait correctement identifié les énoncés qu'il avait lui-même produits.

1. INTRODUCTION

The present study focused on the role of acoustic evidence in a recent criminal investigation regarding the trafficking of narcotics. Under court order, a suspect's telephone line was tapped, and numerous telephone conversations were recorded. Based on perceptual judgements, a police analyst identified another person as being the specific individual in conversation with this suspect, during a number of telephone conversations, including several that were potentially incriminating. This person acknowledged participating in some calls, but denied that he was the speaker in most of the calls, contending that these other calls had been made by another individual. In order to provide evidence for this claim, the defense lawyer contacted the second author (D.G.J.) to request a voice analysis. The present study was undertaken to determine the likelihood that the police hypothesis was correct, and that the individual was in fact the speaker on all the recorded calls, including those that were incriminating.

1.1 Issues in Speaker Identification

Two basic approaches to speaker identification are distinguished: "technical" and "naive" (Nolan, 1983). The first category includes all methods that involve informed analysis, such as visual comparison of spectrograms, acoustic analyses (e.g., comparison of fundamental frequency, formant frequency, formant bandwidth, etc.), and comparison via phonetically trained listening. The second category includes those methods in which no specialized training is required for analysis. The identification is achieved behaviourally (through perceptual judgements), with listeners who are not phonetically trained. "Earwitness" testimony falls into this category.

In the present case, the quality of the recordings complicated definitive acoustic analysis, as many of the telephone calls had been made from cellular telephones, and the spectral characteristics of the signal and the noise varied widely from call to call. Visual spectrographic inspection was also rejected, since this method has been shown to be less accurate than naive perceptual judgement (Nolan, 1983). In view of these limitations, the current study employed naive behavioural speaker identification[1].

Legal evidence based on naive behavioural speaker identification (earwitness testimony), has been surrounded by controversy through its entire history of usage (McGehee, 1937). The legal value of such evidence remains in question (Van Wallendael, Surace, Parsons & Brown, 1994). Earwitness judgments are subject to the effects of expectancy bias (Orchard & Yarmey, 1995), limited attention (Armstrong & McKelvie, 1996; Hammersley & Read, 1985; Saslove & Yarmey, 1980), and memory inaccuracy (Clifford, Rathborn & Bull, 1981; McGehee, 1937; Palmeri, Goldinger & Pisoni, 1993; Saslove & Yarmey, 1980). Interestingly, the psychological literature is replete with studies expressing similar limitations and failings with respect to eyewitness testimony (e.g., Lipscomb, McAllister & Bregman, 1985; Rantzen & Roslyn, 1992), although eyewitness testimony remains an important and valued form of legal evidence.

In part, concerns over earwitness testimony stem from the fact that the testimony is based on acoustic events, which are transient and intangible. While an eyewitness can usually identify a perpetrator with little difficulty, an earwitness has the difficult task of inferring the identity of a perpetrator on the basis of voice information (Hollien et al., 1983; Legge et al. 1984). One aspect of the challenge of speaker identification is the absence of any means of generating a comparative stimulus from witness descriptions. While forensic sketch artists are often able to generate a picture of a perpetrator on the basis of a witness description, they are not similarly able to generate a sample of a perpetrator's voice. Voice information lacks a static representation which can be easily drawn or described, and thus speaker identification is extremely vulnerable to error (Hollien, 1990; Saslove & Yarmey, 1980). Nevertheless, speaker identification is necessary for valuable earwitness testimony, just as visual identification is necessary for eyewitness testimony. This is a negative implication for earwitness testimony, and contributes to its uncertain status. However, regardless of the difficulties with speaker identification, earwitness testimony relating to a recorded or remembered voice may constitute critical evidence in a case.

Given that speaker identification requires the use of recorded voice samples, an important consideration in such identification tasks concerns the length of the voice sample required for accurate identification. Research indicates that sample duration does affect identification performance, but not in a straightforward manner. Yarmey and Matthys (1992) found that hit rate did not reliably increase, and false alarm rate did not reliably decrease as voice sample duration increased from 18 seconds to six minutes. However, in other work, hit rate was reported to increase when voice sample duration was increased from 30 seconds to eight minutes

1. An alternative, using phonetically trained listeners, lacks supportive research (Nolan, 1983), and its potential superiority over naive speaker identification has not been established. Also, unlike naive identification, the influence of various factors on phonetically-based identification is unknown. Moreover, research has demonstrated that listeners are able to identify speakers accurately without any special training (Armstrong & McKelvie, 1996; Bull, Rathborn & Clifford, 1984; Goggin, Thompson, Strube & Simental, 1991; Hollien, Bennet & Gelfer, 1983; Kreiman & Papcun, 1991; Legge, Grosmann & Pieper, 1984; Palmeri et al., 1993; Yarmey, 1994).

(Orchard & Yarmey, 1995). In contrast, smaller increases in sample duration have not reliably resulted in superior speaker identification (Bull & Clifford, 1984). In one study (Haggard, & Summerfield, 1982, cited in Bull & Clifford, 1984) speech samples of less than two seconds produced poor recognition accuracy. It has been shown, however, that speaker identification can be performed with high accuracy rates on the basis of a single syllable. Bricker and Pruzansky (1966) demonstrated that naive listeners were 84% accurate when identifying familiar speakers on the basis of a syllable. Similarly, Williams (1964, cited in Bull & Clifford, 1984) found that listeners could identify a speaker with 93% accuracy in a same-different task with one-syllable utterances, although error rates were lower for two and three-syllable utterances. Pollack, Pickett, and Sumby (1954) found that recognition accuracy for voice improved little with increases in duration beyond one second. While the literature is by no means unanimous, we concluded that identification based on a single word had good prospects for success.

In the current experiment, each voice sample was a token of the word 'okay', extracted from one of the recorded telephone conversations. This selection was motivated by the assumption that speaker identification would be facilitated in a text-dependent context (i.e. with phonemically identical samples), and 'okay' was the largest common phonemic element across the utterances. Comparing (potentially) different voices within the structure of a single word permitted direct phonemic contrasts (e.g., comparing /k/ to /k/) as opposed to non-phonemic contrasts (e.g., comparing /k/ to /n/). Moreover, in accordance with research demonstrating successful speaker identification with one syllable (Bricker & Pruzansky, 1966; Williams, 1964, cited in Bull & Clifford, 1984), the length of the sample was deemed sufficient.

1.2 Objectives

The objective of the study was to test the police hypothesis that the suspect had spoken on each of a particular set of calls. This hypothesis was tested by determining the likelihood of there being more than one voice in the set of attributed utterances, on the basis of same/different listener ratings. A high likelihood of there being more than one voice in the set would constitute evidence against the police hypothesis. Conversely, given that the individual had admitted to producing some of the utterances, a high likelihood of there being only one voice in the set would constitute evidence in favour of the police hypothesis.

Subsequent evaluation of the police hypothesis was accomplished in relation to the suspect's contention. Prior to the experiment, an interview was conducted in which the suspect identified all of the calls in which he had participated. On this basis, the utterances were divided into two categories: utterances from calls in which the suspect admitted participating, and utterances from calls in which the suspect denied participating. Mean similarity ratings were then determined for calls within and between these categories. Higher similarity ratings for calls within categories than between categories would support the suspect's contention, and provide evidence against the police hypothesis. Lack of any differences in similarity ratings would support the police hypothesis.

2. METHOD

2.1 Stimuli

Twenty-four samples of the word 'okay' were selected from the utterances recorded by the police. These samples were taken from 16 separate telephone calls, and thus represented 16 potential speakers. The distribution of the 24 samples was as follows; 10 of the calls contained a single sample, 5 calls contained two samples, and 1call contained four samples. An additional 16 samples were obtained during a subsequent telephone conversation with the suspect, and recorded to tape with the suspect's full consent. In total, 40 samples were extracted from 17 telephone calls, and thus 17 potential speakers were represented in these samples. Samples were selected from calls in which the suspect admitted participating and from calls in which the suspect denied participating.

All conversations had initially been recorded on audio cassette tape. The utterances were digitized with 16-bit resolution at a frequency of 22 kHz, low- pass filtered at 10 kHz., and edited using CSRE (Avaaz, 1996). Editing isolated the word 'okay' from the surrounding acoustic information, and saved each utterance to an audiodata format (.adf, Avaaz, 1996) file.

2.2 Procedure

Two subjects who reported normal hearing ability rated pairs of samples of 'okay' as same or different (i.e., same or different speaker), and indicated the certainty of their response. There were four response alternatives: same-certain, same-uncertain, different-certain, and different-uncertain. Subjects were informed as the nature of the task, and were aware that the number of voices present in the sample was unknown.

Samples were presented to the raters monaurally, at a comfortable listening level. All samples were presented via ER-3A insert earphones, using the ECoS/Win experiment controller (Avaaz, 1997). Samples in each pair were played successively, and could be repeated as many times as desired by the listener. After each response, the computer recorded the response, and cycled to the next trial. The next pair was presented automatically, following a 500 ms interval.

All possible pairs of the 40 stimuli were used in the task, except that no stimulus was ever paired with itself. Thus, each listener rated 1560 pairs of the word 'okay' (780 pairs in both orders). The pairs were randomized across 20 blocks, containing 78 pairs each. Each listener completed the blocks in random order, with a short rest between each block. After completing 10 blocks (780 pairs), the test session stopped, and the remaining 10 blocks were completed on a subsequent day. Subjects required approximately one hour to complete each set of 10 blocks.

3. RESULTS AND DISCUSSION

3.1 Rater Accuracy

An estimate of rater accuracy was generated by comparing the hit rates for pairs in which both samples had been extracted from a single telephone call, as these pairs had to have been produced by the same speaker. Responses of "same" for such pairs were counted as hits without respect to certainty. The samples obtained in our own recordings were not included in this comparison, however, because the quality of these samples was superior to that of the samples obtained by the police. Overall, the average hit rate was 0.89. This estimate is similar to that reported by Williams (1964, cited in Bull & Clifford, 1984), who found that subjects could identify a speaker with 93% accuracy in a same-different task with only one-syllable. This estimate of accuracy should be treated with caution, however, because samples from within a telephone call also shared acoustic information apart from voice spectra (such as specific telephone noise), which could have contributed to the same-different decision, and inflated the accuracy rate.

3.2 Analysis of Variance

The primary goal of the study was to test the police hypothesis that there was but one speaker in the set of imputed utterances. For the purposes of analysis, the utterances were grouped according to the telephone calls from which they had been extracted, such that all utterances extracted from a call were considered to be equivalent (i.e., produced by the same speaker). Since the utterances from the police recordings were extracted from 16 different telephone conversations, 16 groups of utterances were formed, representing 16 potential speakers. Utterance pairs were accordingly recoded as 126 pairs[2] of potential speakers. The utterance pair ratings were then subjected to an ANOVA with potential speaker pair and rater as factors.

The effect of potential speaker pair was significant ($F(125{,}251) = 14.15, p < 0.001$), indicating that the ratings (i.e., same-certain, same-uncertain, different-certain, & different-uncertain) differed significantly with potential speaker pair. This result indicates that there were highly reliable differences in the judged similarity of utterances across recordings, which suggests that the police hypothesis was incorrect, and that the utterances may have involved more than one speaker. However, significant differences in ratings across potential speaker pairs could also have reflected different levels of certainty in the ratings, which would be expected, given the varying quality of the samples obtained by the police. Thus, the data were recoded to collapse across levels of certainty, and a second ANOVA was conducted. The effect of potential speaker pair was highly significant in this analysis ($F(125{,}251) = 10.80, p < 0.001$), indicating that the significant differences in the same/different ratings could not be attributed to differences in certainty. Moreover, given that the original police hypothesis had also been generated by listening to voice samples of varying quality obtained in the telephone tap, the different conclusion reached by these raters cannot be easily dismissed, and provides sufficient reason to doubt the police hypothesis.

There was also a significant main effect of rater ($F(1{,}125) = 4.05, p < 0.05$) and a significant interaction between potential speaker pair and rater ($F(125{,}251) = 1.50, p < 0.005$). This indicates that the two raters were not in complete agreement, or did not share the same degree of certainty regarding their decisions. To evaluate these possibilities, the effect of rater was examined with data recoded to collapse across levels of certainty. There was no significant main effect of rater in this re-analysis, suggesting that the differences between the ratings made by the different raters were based on differences in certainty. However the interaction between rater and call pair was significant ($F(125{,}251) = 1.32, p < 0.05$), indicating that the raters may have used different factors or weights in their rating decisions[3].

For the next analysis, samples were categorized in accordance with whether they had been extracted from a call in which the suspect admitted participating (A), denied participating (D), or had participated in for the purposes of this experiment (S). On the basis of these types, each potential speaker pair was categorized as one of six possible combinations. The ratings

2. There were only 120 pairs of potentially different speakers. The six remaining pairs were instances where multiple utterances were extracted from the same call.

3. Since short (two-syllable) samples were used in the present study, it might be suggested that the samples were not long enough for the raters to make accurate voice similarity ratings, which would discount the decisions of these raters in relation to the decisions of the police. However, because the raters in the present study were reasonably accurate in their decisions (e.g., hit rate of 0.89), the lack of complete agreement between them does not undermine their challenge to the police hypothesis.

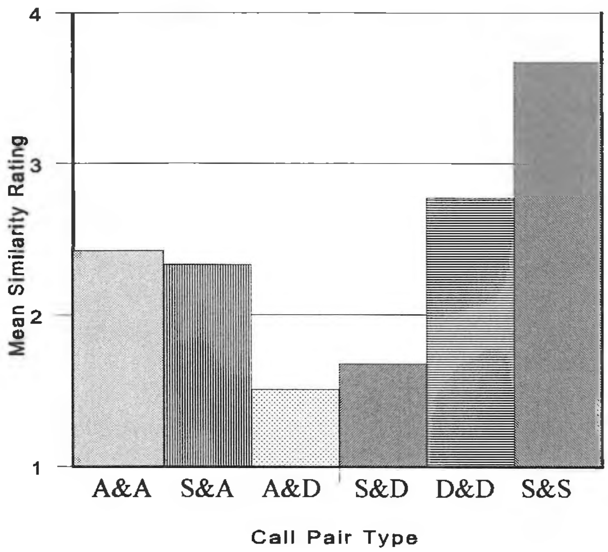


Figure 1: Similarity ratings for pairs of utterances from calls in which the suspect participated (S), admitted participating (A), or denied participating (D).

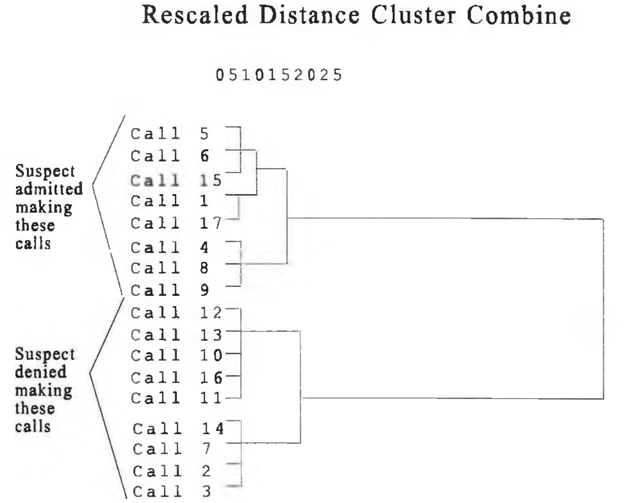


Figure 2: Results of cluster analysis on similarity ratings for all calls

were then subjected to a one-way ANOVA with speaker pair type as the independent variable. The effect of speaker pair type was significant (F (5,3119) = 666.07, $p < 0.001$), indicating that similarity ratings were different for the various speaker pair types. Mean similarity ratings were calculated for each type of pair[4], and are presented in Figure 1. A Tukey's HSD (Honestly Significant Difference) posthoc test of means indicated that all means were significantly different, except for the means of the first call pair type (A & A), and the second call pair type (S & A). The highest mean similarity ratings were obtained for pairs in which both samples were extracted from the recordings of the suspect obtained by our lab. This is not surprising, as these samples were extracted from a single telephone conversation (thus reflecting voice characteristics at a single point in time), were of relatively high quality, and were free from background and channel noise. Pairs of samples extracted from calls in which the suspect admitted participating were rated as fairly similar to each other, as were pairs involving these samples and samples from our recording of the suspect. However, both of these types of samples were rated as less similar to samples from calls in which the suspect denied participating. Conversely, samples from calls in which the suspect denied participating were rated as highly similar to each other. These results suggest that the suspect had been accurate in identifying calls in which he had and had not participated, and that the police hypothesis is incorrect.

4. A rating of four was equivalent to the decision that samples were certainly from the same speaker (i.e., same-certain). A rating of one was equivalent to the decision that samples were certainly from different speakers (i.e., different-certain).

The high similarity ratings for samples extracted from calls in which the suspect denied participating suggest that these samples likely were produced by another single speaker, and accordingly support the suspect's initial contention that most of these calls were made by another single speaker. Nevertheless, it is also possible that these calls were made by the suspect using a disguised voice, or a number of speakers with highly similar voices. While the latter possibility is unlikely, it is difficult to rule out the possibility that the suspect successfully disguised his voice.

3.3 Cluster Analysis

In order to gain a better understanding of the relationships between the samples, a cluster analysis was performed, using Ward's method of agglomeration (Ward, 1963). The analysis was performed on a matrix in which each of the 17 calls was treated as both a distinct case and a unique variable. The value for each of the cases on each of the variables was accordingly defined as the mean similarity rating for that call pair[5]. Clusters were thus formed on the basis of perceived voice similarity, as coded by the ratings of the unique pairs[6].

5. Prior to computing the mean similarity ratings for each call pair for the cluster analysis, the data was recoded such that higher similarity ratings were represented by lower numbers. Thus, for this analysis, a rating of one was equivalent to the decision that samples were certainly from the same speaker, and a rating of four was equivalent to the decision that samples were certainly from different speakers. Mean similarity ratings could then be treated as distances between call pairs.

6. Using Ward's method of agglomeration, cases or clusters are combined in sequential order based on squared Euclidean

The samples extracted from our own recording were included in this analysis, in order to determine the perceptual relationship between the suspect's voice and the other voice samples. Thus, 143 pairs of potential speakers, based on 17 different calls, were subjected to the cluster analysis. From this analysis, two distinct clusters emerged (see figure 2). These distinct clusters clearly suggested the presence of two strong voice percepts in the set of samples, and supported the suspect's contention that the police hypothesis was incorrect.

The clusters could also be explained, however, on the basis of telephone channel noise or spectral distortion resulting from the use of certain telephones, or background noise specific to particular calling locations. It could be argued that all of the calls had been made by a single speaker, but the voice characteristics on the calls in each cluster differed due to the use of different cellular telephones, each effecting a particular distortion of the voice and adding specific channel noise. While this argument is tenable, it seems more likely that the clusters represent two different voices. The suspect's voice, obtained in a staged telephone tap for the purposes of the experiment, fell neatly into one of the clusters. Unlike the 16 calls recorded by the police, this call (coded as call number 17- see figure 2) had very little channel noise, and very little spectral distortion. If the clusters were based on similarities of channel noise and spectral distortion, call number 17 would not have clustered with any of the police calls. The fact that this call fell into a cluster strongly suggests that the clusters were not based on telephone channel noise and spectral distortion, but rather on voice characteristics.

Another feature of the analysis provides support for the suspect's contention. When the results of the cluster analysis are compared with the suspect's claims regarding his participation in each of the calls, an interesting pattern emerges. Every call in which the suspect admitted participating (1, 4, 5, 6, 8, 9 & 15) fell into the first cluster, and every call in which the suspect denied participating (2, 3, 7, 10, 11, 12, 13, 14, & 16) fell into the second cluster. Moreover, the recording of the suspect's voice (call 17) fell into the first cluster (i.e., with all the other calls to which he admitted). This is strong evidence in support of the suspect's testimony, and in opposition to the police hypothesis.

4. GENERAL DISCUSSION

The results of the experiment suggest that the suspect did not produce all of the utterances attributed to him by the police. The analysis of variance indicates that similarity ratings were not uniform for all paired-comparisons. And while the raters in this experiment were not in total agreement, the relatively high degree of accuracy that they achieved indicates that their decisions cannot be easily dismissed in favour of the decision of the police. Most importantly, the results of the cluster analysis clearly indicate the presence of more than one voice in the set of utterances attributed to the suspect. The cluster analysis also supported the suspect's claims in regard to the specific calls in which he did and did not participate. Thus, the present experiment provides numerous lines of evidence in support of the suspect's contention that the police hypothesis was incorrect.

As is common in forensic investigations, the results of the present study do not provide conclusive evidence regarding the innocence or guilt of the suspect. In fact, it is quite possible that the all of the utterances were produced by the same speaker, in spite of the evidence to the contrary. For example, there remains the possibility that the suspect disguised his voice on certain calls (e.g., calls relating to criminal activity, which he would later not admit to making), intentionally creating a false voice percept. Such a disguise, if performed convincingly, could have produced the results of the present study, given that listeners may have difficulty distinguishing a disguised voice from a different voice (Hollien, 1990). In the present case, the likelihood of attempted voice disguise is supported by evidence that speakers suspected that their telephone conversations were being monitored, in that the speakers discussed potentially incriminating matters only indirectly. Thus, although the results of the study suggest that the police hypothesis is incorrect, they clearly do not establish the innocence or guilt of the suspect.

However, the value of such research can be found in the various ways in which it improves the process of forensic investigation. This study sought to quantify the degree to which the voice samples were similar or different, via paired-comparison ratings. In contrast, the police hypothesis was based on the conjecture of officers transcribing audio tapes, and the likelihood of its veracity could not be established quantitatively. This has important implications for the admission of such evidence in the legal system. The value of any evidence in the legal system must be weighed carefully according to the likelihood that the evidence is accurate. Thus, providing statistics which can reasonably quantify the value of evidence may assist the legal process.

REFERENCES

Armstrong, H.A. & McKelvie, S. J. (1996) Effect of face context on recognition memory for voices. *J. of General Psychology, 123 (3)*, 259-270.

Avaaz Innovations. (1996) *Computerized Speech Research Environment (CSRE)*, Version 4.5, Avaaz Innovations Inc., London, Ontario, Canada.

---

distances. Specifically, at each successive step of the analysis, the two clusters or cases for which combination will produce the smallest increase in the sum of squared within-cluster distances are combined.

Avaaz Innovations. (1997) *ECoS/Win: Experiment Generator and Controller,* Avaaz Innovations Inc., London, Ontario, CANADA.

Bricker, P. & Pruzansky, S. (1966) Effects of stimulus content and duration on talker identification. *J. Acoust. Soc. Am., 40(6),* 1441-1450.

Bull, R. & Clifford, B.R. (1984) Earwitness voice recognition accuracy. In G.L. Wells & E.F. Loftus (eds.), *Eyewitness Testimony,* (92-123). New York: Cambridge University Press.

Clifford, B.R., Rathborn, H. & Bull, R. (1981) The effects of delay on voice recognition accuracy. *Law and Human Behavior, 5,* 201-208.

Goggin, J.P., Thompson, C.P., Strube, G. & Simental, L.R. (1991) The role of language familiarity in voice identification. *Memory and Cognition, 19(5),* 448-458.

Hammersley, R. & Read, J.D. (1985) The effect of participation in a conversation on recognition and identification of the speakers' voices. *Law and Human Behavior, 9(1),* 71-81.

Hollien, H. (1990) *The Acoustics of Crime.* New York: Plennum Press.

Hollien, H., Bennet, G. & Gelfer, M.P. (1983) Criminal identification comparison: aural versus visual identification resulting from a simulated crime. *J. of Forensic Sciences, 28(1),* 208-221.

Kreiman, J. & Papcun, G. (1991) Comparing discrimination and recognition of unfamiliar voices. *Speech Communication, 10,* 265-275.

Legge, G.E., Grossman, C. & Pieper, C.M. (1984) Learning unfamiliar voices. *J. of Experimental Psychology: Learning, Memory, and Cognition, 10(2),* 298-303.

Lipscomb, T.J., McAllister, H.A. & Bregman, N.J. (1985) Bias in eyewitness accounts: The effects of question format, delay interval, and stimulus presentation. *Journal of Psychology, 119,* 207-212.

McGehee, F. (1937) The reliability of the identification of the human voice. *J. of General Psychology, 17,* 249-271.

Nolan, F. (1983) *The Phonetic Bases of Speaker Recognition.* New York: Cambridge University Press.

Orchard, T.L. & Yarmey, D.A. (1995) The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology, 9,* 249-60.

Palmeri, T.J., Goldinger, S.D. & Pisoni, D.B. (1993) Episodic encoding of voice attributes and recognition memory for spoken words. *J. of Experimental Psychology: Learning, Memory, & Cognition, 19(2),* 309-328.

Pollack, I., Pickett, J. & Sumby, W. (1954) On the identification of speakers by voice. *J. Acoust. Soc. Am., 26(6),* 403-412.

Rantzen, A. & Roslyn, M. (1992) The reversed eyewitness testimony design: more evidence for source monitoring. *J. of General Psychology, 119(1),* 37-43.

Saslove, H. & Yarmey, A.D. (1980) Long-term auditory memory: speaker identification. *J. of Applied Psychology, 65(1),* 111-116.

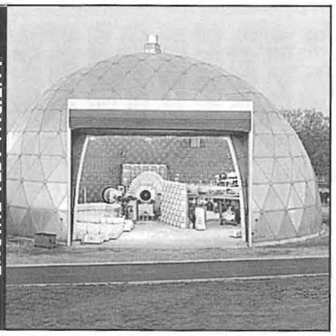
Van Wallendael, L.R., Surace, A., Parsons, D.H. & Brown, M. (1994) Earwitness voice recognition: Factors affecting accuracy and impact on jurors. *Applied Cognitive Psychology, 8,* 661-677.

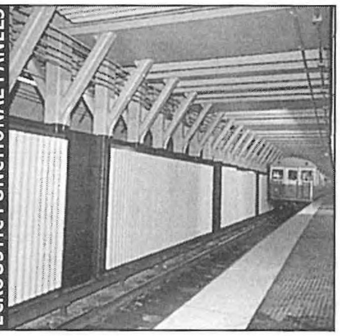
Ward, J.H. (1963) Hierarchical grouping to optimize an objective function. *J. of the American Statistical Association, 58,* 236-244.

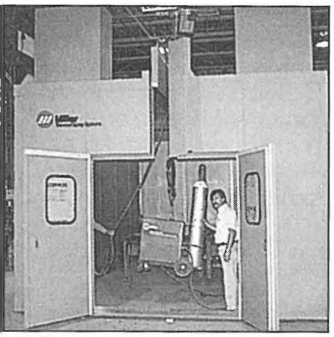
Yarmey, A.D. (1994) Earwitness evidence: Memory for a perpetrator's voice. In D.F. Ross, J.D. Read & M.P. Toglia (eds.), *Adult Eyewitness Testimony: Current Trends and Developments* (101-124). New York: Cambridge University Press.

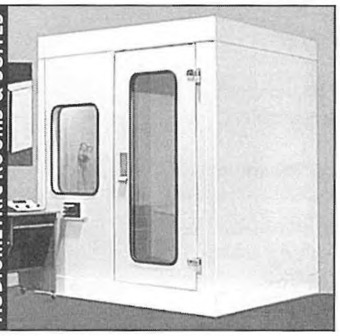
Yarmey, A.D. & Matthys, E. (1992) Voice identification of an abductor. *Applied Cognitive Psychology, 6,* 367-377.

* *Work supported by NSERC grant to D.G. Jamieson. Address correspondence to jamieson@audio.hhcru.uwo.ca.*


ENGINE TEST FACILITY

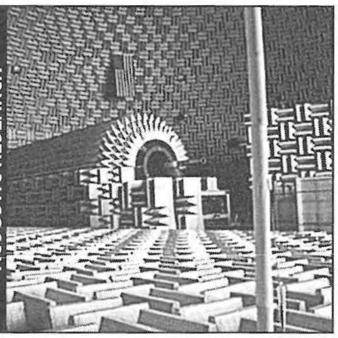

ECKOUSTIC FUNCTIONAL PANELS

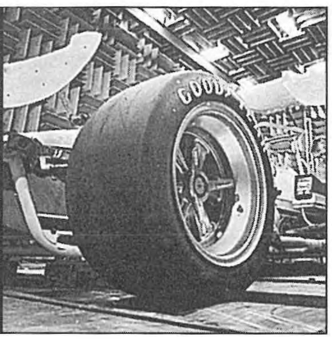

O.E.M. ENCLOSURE

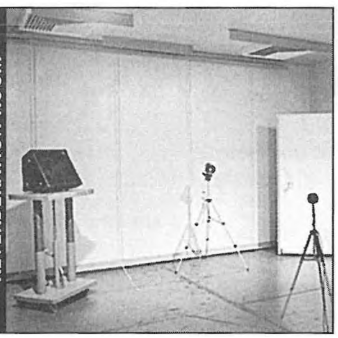

AUDIOMETRIC ROOMS & SUITES

SOUND SOLUTIONS FOR THE FUTURE


ARCHITECTURAL ACOUSTICS


ACOUSTIC RESEARCH


AUTOMOTIVE TEST CHAMBER


REVERBERATION ROOM

ECKEL

NOISE CONTROL TECHNOLOGIES

CANADIAN OFFICE

Box 776 100 Allison Avenue Morrisburg ON K0C 1X0
Tel: 613-543-2967 800-563-3574 Fax: 613-543-4173
Web site: www.eckel.ca/eckel e-mail: eckel@eckel.ca

ECKEL NOISE CONTROL TECHNOLOGIES