The ability of expert witnesses to identify voices: a comparison between trained and untrained listeners

Niels O. Schiller[*] *and Olaf Köster*[†]

[*] *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands and Cognitive Neuropsychology Laboratory, Department of Psychology, Harvard University, Cambridge, MA*
[†] *Bundeskriminalamt (BKA), Wiesbaden, Germany*

ABSTRACT This study reports the results of a speaker identification experiment in which the performance of phonetic expert witnesses and untrained listeners was compared. In a direct identification task participants from both groups were asked to identify the voice of a target speaker among five foils. Results showed that expert witnesses, who were experienced in speaker identification, performed significantly better than untrained listeners, who had no experience in phonetic speaker identification.

KEYWORDS speaker identification, phonetic expert witness, forensic phonetics

INTRODUCTION

Both acoustic and linguistic information have been shown to play a role in speaker identification (Goldstein *et al.* 1981; Ladefoged and Ladefoged 1980). Speaker identification is generally improved when untrained listeners have some knowledge of the language of the target speaker as compared to a situation when they do not know his or her language (Goggin *et al.* 1991; Sullivan and Schlichting 1997; Thompson 1987). Recent experimental evidence suggested that familiarity with the target language had a positive effect on speaker identification (Köster and Schiller 1997; Köster, Schiller and Künzel 1995; Schiller and Köster 1996; Schiller, Köster and Duckworth 1997). In a number of experiments participants performed significantly better in a speaker identification task when they had some knowledge of the target language compared to when they did not know the target language. When virtually all linguistic information of the target language was removed from the stimulus materials, untrained listeners differing in native-language background (including the target language) did not statistically differ from each other in identify-

1350–1771
Forensic Linguistics 5(1) 1998

ing the target speaker. This supports the assumption that speaker identification involves the processing of linguistic information.

This assumption is further supported by a recent study by Remez, Fellowes and Rubin (1997). These authors reported a series of experiments that assessed the ability of untrained listeners to identify familiar voices from phonetic attributes alone when presented with speech samples that lacked the acoustic correlates of natural voice quality. By using *sinewave replication*, i.e., an acoustic technique that preserves the phonetic properties of speech while discarding the acoustic attributes of voice quality and intonation, Remez *et al*. (1997) generated sentences that were intelligible but sounded unnatural in timbre. Their results showed that listeners were able to identify talkers by using solely information about linguistically governed articulation without acoustic information about voice quality.

The present study investigates whether phonetically trained expert witnesses are more reliable in speaker identification tasks than untrained listeners. A former study by Köster (1987) suggested that this is in fact the case. Köster (1987) carried out a series of speaker identification experiments comparing phonetic experts with naive listeners. His results indicated that phoneticians performed better than naive listeners. Köster (1987) concluded that phonetic experts were able to make a reliable decision about the identity/non-identity of two voice samples whereas this was not the case for naive (phonetically untrained) listeners.

Phonetically trained expert witnesses have knowledge about acoustic as well as linguistic aspects of the speech signal. Therefore, it may be hypothesized that their ability to discriminate between a target speaker and a number of foils in a speaker identification task should be (significantly) better than the performance of a control group of untrained listeners.

EXPERIMENT

The above-mentioned hypothesis was tested in a speaker identification experiment under laboratory conditions. A group of phonetic expert witnesses (trained listeners) and a control group of phonetically untrained listeners were selected and tested on a speaker identification task using the same stimulus materials.

METHOD

Participants

There was a total of twenty-seven native German listeners divided into two groups. The first group consisted of seventeen phonetically untrained

listeners. All of them were undergraduate students of the University of Trier. In the second group, there were ten phonetic expert witnesses who were experienced in speaker identification. They came from either governmental institutions such as the Forensic Science Laboratories of the German Federal Criminal Bureau or the State Criminal Bureaus or they came from phonetics departments of universities. All participants took part in the experiment voluntarily. None of them reported any hearing problems.

Materials

The speech materials used in the experiment came from six German native speakers. They were recorded using a Sony DAT recorder while reading the following passage in German:

> Guten Tag, hier ist Meier, es geht um folgendes: Wir haben Ihre kleine Tochter Ramona nach der Schule in unsere Obhut genommen. Wenn Sie nicht wollen, daß ihr etwas passiert, dann hören Sie jetzt mal gut zu. Besorgen Sie sich vierzigtausend Mark. Packen Sie das Geld in einen schwarzen Koffer. Sie spielen dann selbst den Boten. Fahren Sie mit ihrem Auto in Richtung Scheef. Bei dem Schuppen an der Ausfahrt Camberg bleiben Sie stehen. Und noch etwas: Lassen Sie die Polizei aus dem Spiel. Wenn Sie den Schuppen erreicht haben, stellen Sie den Motor ab und bleiben im Wagen sitzen. Sie hören dann von uns. Wenn Sie glauben, Sie könnten quer schießen, dann liegen Sie schief, dann kriegen Sie keinen Fuß mehr auf die Erde. Wir werden Sie auf gar keinen Fall schonen, und das wäre doch sehr schade.

The passage was approximately one minute in length when read aloud. From each of the six speakers three parts of the passage were spliced out of the recordings using a wave form editor (Computerized Speech Lab, Kay Elemetrics Corporation), each between four and eight seconds in length. To obtain exactly the same materials under telephone transmission conditions,[1] the three speech samples from each speaker were recorded again through an analogue telephone line yielding a total of six samples from each speaker. Each of these six samples was re-recorded three times so that finally there were 108 speech samples. One speaker was chosen to be the target, the other five were foils.

Procedure and design

The two groups were tested separately. Listeners were first familiarized with the voice of the target speaker by listening five times to the whole text passage. Familiarization took about five minutes. Listeners were instructed to memorize the voice of the target speaker as accurately as

possible. After a short break of approximately five minutes, listeners were exposed to a forced-choice test. They listened to a test tape containing the 108 speech samples in a randomized order. Their task was to mark 'yes' on their response sheets whenever they recognized a speech sample as coming from the target speaker. They marked 'no' if they thought a speech sample came from one of the foils. The entire voice line-up was presented only once and had a duration of approximately thirty minutes.

RESULTS

Discrimination sensitivity, i.e., the ability to discriminate between targets and foils, was measured by *d'* from Signal Detection Theory (Macmillan and Creelman 1991). *d'* is a specific measure of the discrepancy between a hit rate (*H*) – i.e., the proportion of target trials to which the participants responded 'yes' – and a false-alarm rate (*F*) – i.e., the proportion of foil trials to which the participants (incorrectly) responded 'yes' (for a more detailed description of the analysis see Schiller, Köster and Duckworth 1997).

Hits and false alarms were summarized for each participant and then pooled across groups. Since the overall recognition rate was quite high and there were no theoretically interesting differences between the two groups when high fidelity and telephone transmission trials were analysed separately, all trials were collapsed for the analysis.

Group 1 (17 untrained listeners) yielded 282 hits out of 306 target voice trials and 37 false alarms out of 1530 foil trials (see Table 1). This equals a hit rate of 0.92 and a false alarm rate of 0.02.

Table 1 Distribution of responses of group 1 (untrained listeners, n = 17)

	response		
stimulus class	'yes'	'no'	total
target voice	282 (H = 0.92)	24	306
dummy voice	37 (F = 0.02)	1493	1530

Group 2 (10 expert witnesses) made 177 hits out of 180 target voice trials and 10 false alarms out of 900 foil trials (see Table 2). This corresponds to a hit rate of 0.98 and a false-alarm rate of 0.01.

Table 2 Distribution of responses of group 2 (expert witnesses, n = 10)

	response		
stimulus class	'yes'	'no'	total
target voice	177 (H = 0.98)	3	180
dummy voice	10 (F = 0.01)	890	900

The corresponding *d'* values are 3.46 for group 1 and 4.38 for group 2 (see Figure 1). A statistical comparison revealed that the difference between the two *d'* values was significant ($p<.05$), i.e., the two groups differed significantly from each other with respect to identifying the target speaker.

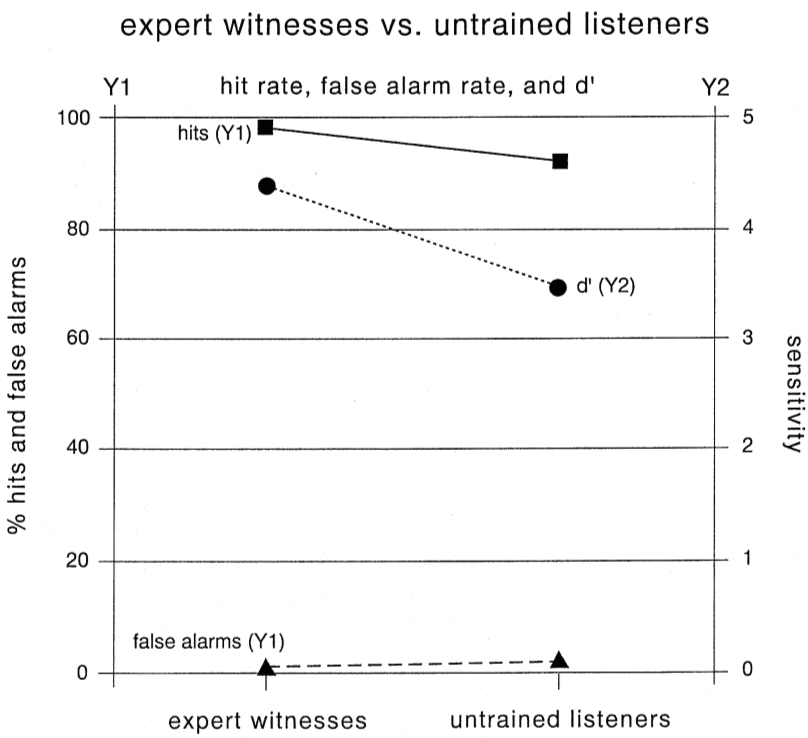

Figure 1 Hit rate (H), false-alarm rate (F), and sensitivity (*d'*) for the two listener groups

DISCUSSION

The hypothesis tested in the experiment reported in this paper was confirmed. If phonetically trained and untrained listeners were exposed to the same speech materials, the phonetically trained listeners performed significantly better in identifying a speaker than the untrained listeners. Although the discrimination sensitivity of the untrained listeners was relatively high compared to untrained listeners with a different native-language background (Köster *et al*., 1995; Köster and Schiller 1997; Schiller and Köster 1996), the phonetic expert witnesses performed still better. More specifically, the difference in discrimination sensitivity between the two groups was statistically significant. In fact, expert witnesses performed at ceiling level.

A methodological problem that has to be discussed here is the repeated presentation of the target voice samples. Bastiaansen *et al*. (1996) have pointed out that the repeated presentation of voices in a voice line-up may be problematic since listeners may identify the voice of a foil as that of the target (false-alarm) and repeatedly select samples from that speaker in the following presentations because these are perceived to be similar. That is, the individual judgements may not be independent of each other and may therefore not be taken to add extra weight to the outcome of the test. Bastiaansen *et al*. designed an experiment in which a sound-alike was used in the line-up instead of the target voice in one condition, i.e., the target speaker was no longer present. Bastiaansen *et al*. (1996) hypothesized that the identification of the sound-alike as the target would be above chance. However, since the sound-alike was not optimally chosen, their results remain somewhat inconclusive.

Recently, however, Sullivan and Schlichting (1997) published a study in which they used speech materials from a professional voice imitator who imitated the voice of their target speaker (see also Schlichting and Sullivan forthcoming). In a speaker identification experiment participants falsely recognized the voice of the imitator as that of the target speaker in up to 100 per cent of the cases. None of their four test groups was able to accurately detect the absence of the target voice in the line-ups. This result adds plausibility to the argument of Bastiaansen *et al*. (1996). However, the potential weakness of repeated presentation of voices in a line-up still has to be proven experimentally. Broeders, Rietveld, and Schiller (personal communication) plan a study to test this issue.

CONCLUSION

The results showed that there were differences between trained and untrained listeners in speaker identification. Phonetically trained native

German expert witnesses performed significantly better than untrained native German listeners. The forensic relevance of this result is straightforward. Forensic situations in which the perpetrator's voice is the only definite piece of evidence make earwitness testimony necessary. However, the reliability of non-expert earwitnesses has been questioned (see Clifford 1980 and Deffenbacher *et al*. 1989 for reviews). Clifford (1980), for instance, concludes from his review 'that the criminal justice system must exercise the greatest caution when utilizing voice identification in either case building or case prosecution' (Clifford 1980: 390). The conclusion by Deffenbacher *et al*. is similarly sceptical: 'earwitnessing is so error prone as to suggest that no case should be prosecuted solely on identification evidence involving an unfamiliar voice' (Deffenbacher *et al*. 1989: 118).

The results of our experiment showed, however, that phonetic expert witnesses are in general more reliable in identifying voices than naive listeners. Forensic phoneticians have the knowledge to examine a speech sample aural-perceptually and to carry out acoustic analyses of features like fundamental frequency or vowel formants. They will look for speaker-specific features in articulation, voice quality, intonation, or respiration. Although there is substantial disagreement as to the value of expert witnesses (see Hollien, 1990 for a review), the present result underlines the importance of expert witnesses for courtroom testimony (Levi, 1994). Therefore, they should be consulted when the identity of voice is in question in a court trial and a recording from the perpetrator is available. Relying on the judgements of untrained earwitnesses should be avoided. Although some layperson listeners may have a high discrimination sensitivity, it is the generally higher false-alarm rate of the untrained listeners which may pose important problems in court trials, namely when an innocent is falsely accused.

ACKNOWLEDGEMENTS

The authors would like to thank the members of the German 'Arbeitsgemeinschaft Sprechererkennung' for their participation in the experiment. The research reported in this paper was supported by a grant from the International Association for Forensic Phonetics (IAFP) to Olaf Köster and Niels O. Schiller. An oral version of this paper was presented at the 1997 Annual Meeting of the IAFP in Edinburgh, Scotland, 6–10 July 1997.

NOTES

1 Although the experiment was carried out under laboratory conditions, we wanted to keep it as close to reality as possible. Therefore, the telephone transmission condition was included since in forensic cases the voice of a suspect has often been transmitted over a telephone line (e.g., obscene phone calls, bomb hoaxes, ransom demands, etc.) before being recorded.

REFERENCES

Bastiaansen, M., Rietveld, A. C. M., and Broeders, A. P. A. (1996) 'Repeated testing in auditory identification by witnesses', paper given at the 1996 Annual Meeting of the International Association for Forensic Phonetics in Wiesbaden (Germany), 7–11 July, 1996.

Clifford, B. R. (1980) 'Voice identification by human listeners: On earwitness reliability', *Law and Human Behavior*, 4: 373–94.

Deffenbacher, K. A., Cross, J. F., Handkins, R. E., Chance, J. E., Goldstein, A. G., Hammersley, R. and Read, J. D. (1989) 'Relevance of voice identification research to criteria for evaluating reliability of an identification', *The Journal of Psychology*, 123: 109–19.

Goggin, J. P., Thompson, C. P., Strube, G., and Simental, L. R. (1991) 'The role of language familiarity in voice identification', *Memory and Cognition*, 19: 448–58.

Goldstein, A. G., Knight, P., Bailis, K., and Conover, J. (1981) 'Recognition memory for accented and unaccented voices', *Bulletin of the Psychonomic Society*, 17: 217–20.

Hollien, H. (1990) 'The phonetician as an expert witness. Ethics and responsibilities', *Annals of the New York Academy of Sciences*, 606: 33–45.

Ladefoged P. and Ladefoged J. (1980) 'The ability of listeners to identify voices', UCLA Working Papers in Phonetics, 49: 43–51.

Levi, J. N. (1994) 'Language as evidence: The linguist as expert witness in North America', *Forensic Linguistics*, 1: 1–26.

Köster, J. P. (1987) 'Auditive Sprechererkennung bei Experten und Naiven', in R. Weiss (ed), *Festschrift für Hans-Heinrich Wängler*, Hamburg: Buske, 171–9.

Köster, O. and Schiller, N. O. (1997) 'Different influences of the native language of a listener on speaker recognition', *Forensic Linguistics*, 4: 18–28.

Köster, O. Schiller, N. O., and Künzel, H. J. (1995) 'The influence of native-language background on speaker recognition', in K. Elenius and P. Branderud (eds), *Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, vol. 3*, Stockholm: KTH and Stockholm University, 306–9.

Macmillan, N. A. and Creelman, C. D. (1991) *Detection Theory: A User's Guide*, Cambridge: Cambridge University Press.

Remez, R. E., Fellowes, J. M., and Rubin, P. E. (1997) 'Talker identification based on phonetic information', *Journal of Experimental Psychology: Human Perception and Performance*, 23: 651–66.

Schiller, N. O. and Köster, O. (1996) 'Evaluation of a foreign speaker in forensic phonetics: A report', *Forensic Linguistics*, 3 (1): 176–85.

Schiller, N. O., Köster, O. and Duckworth, M. (1997) 'The effect of removing linguistic information upon identifying speakers of a foreign language', *Forensic Linguistics*, 4 (1): 1–17.

Schlichting, F. and Sullivan, K. P. H. (1997) 'The imitated voice – a problem for voice line-ups?', *Forensic Linguistics*, 4 (1): 148–65.

Sullivan, K. P. H. and Schlichting, F. (1997) 'Speaker identification in a second language by university-level language students', *Phonum. Reports from the Department of Phonetics, Umeå University*, 4: 129–32.

Thompson, C. P. (1987) 'A language effect in voice identification', *Applied Cognitive Psychology*, 1: 121–31.