DISCRIMINATION OF IMITATED VOICES

F. Schlichting and K.P.H. Sullivan

Phonetics Section
Department of Linguistics
Umeå University, Sweden.

ABSTRACT - The belief that an individual can discriminate between different voices from memory is central to the concept of the voice line-up and its application in the legal sphere. The question of how voice imitation can affect the accuracy of a voice line-up has received little attention. However, recent research has demonstrated that high-quality imitation can cause a problem for the reliability of speaker discrimination within a voice line-up. This paper examines the effect of amateur imitations on the reliability of speaker discrimination within in the line-up. The voice which was imitated and the line-up construction was the same as in the study using the professional imitator. Unlike in the study which used the professional imitation, there was almost no confusion between the imitation and the real voice — the listeners were not convinced that the voice they were to recognize was someone else they knew. However, when the imitations were absent from the line-ups there was some confusion between the natural voice and the not present option. In fact, the amateur imitators, even though they failed to cause confusion with the voice they were imitating, succeeded in disguising their voices to cause a high number of judicially fatal errors.

INTRODUCTION

Adult listeners are particularly adept at recognizing people by their voices alone. Interestingly this skill extends beyond close friends and family, those individuals people often associate with daily on a face-to-face basis, to public figures such as television personalities and national politicians. The ability to recognize someone purely aurally is widely reported in the literature on speaker recognition (eg. Künzel, 1987) and forms one of the central pillars of the voice line-up. This identification technique can be viewed as the aural version of the police identification parade. The technique is appropriate to several criminal situations which include those in which the only piece of tangible evidence available to the police or the prosecution is the perpetrator's voice.

The ability to discriminate between voices is a second skill which is crucial for voice line-ups to be a useful forensic tool. Rose & Duncan (1995) investigated the ability of listeners to identify and discriminate between individuals with similar voices whom they knew well, ie. the listeners were all close family friends or relatives to the individuals with similar voices. Rose & Duncan conducted a range of small-scale perception experiments. These included identification tests in which the listeners had to identify the speaker or write 'don't know' and discrimination tests where the listeners had to respond 'yes/no' to same/different pairs of stimuli. Their findings showed that the incidence of judicially fatal errors in identification and discrimination ranged from 18 per cent for single word utterances to 15 per cent for utterances of about 45 seconds in length. It is, therefore, important to assess the degree of inaccuracy which can occur in a voice line-up due to a similar voice.

Recently, Schlichting (1996) and Schlichting & Sullivan (forthcoming) have extended Rose & Duncan's work on the similar voice by investigating the impact of a professional imitation on voice line-ups. Schlichting & Sullivan used an imitation of a well-known Swedish politician, the former *Statsminister,* Carl Bildt, provided by Göran Gabrielsson in their study. Göran Gabrielsson was not personally known to any of the listeners performing the voice line-up identification tasks. In their studies Carl Bildt was the familiar voice, which together with the imitation of Carl Bildt by Göran Gabrielsson formed a similar pair of familiar voices.

Imitation, here and in the Schlichting (1996) and Schlichting & Sullivan (forthcoming) papers, is defined as when the speaker not only attempts to disguise their voice so that it is not recognized (c.f. speaker disguise (Künzel, 1987)), but also attempts to manipulate their voice in such a manner that it will be mistaken as the voice of another person. This initial work using a high-quality professional imitation showed that when the imitation and the natural voice of the imitator are both absent from the

line-up the listeners almost universally choose the person whose voice has been imitated regardless of whether the voice to be identified was that voice or the imitation of that voice. Interesting, though, the work also demonstrated that the listeners were able to descriminate between the natural voice and the imitated version of the same voice.

This paper describes several experiments, which continue Schlichting & Sullivan's (forthcoming) examination of the imitated voice and the problems it can cause for line-up accuracy. The situation examined in the paper can be considered a more realistic one from the forensic viewpoint. The imitations used for the experiments presented here are amateur imitators. Good imitation is a skill which is difficult to acquire and thus it is unlikely that a criminal would be able to imitate another individual as successfully as Göran Gabrielsson imitated Carl Bildt in Schlichting & Sullivan's earlier work. This hypothesis is supported by the low level of confusion in the earlier studies with the amateur imitations which were used as foils in the line-ups.

In order to confirm this hypothesis one of the earlier experiments was repeated twice with the professional imitation replaced by an amateur imitation and the natural voice of the professional imitator replaced with the natural voice of the amateur imitator.

PROCEDURE

Subjects: speakers

The two sets of voice line-ups used in this study were constructed by taking the line-ups used in the earlier studies and exchanging the professional imitation with the new amateur imitations and the voice of Göran Gabrielsson with the natural voices of the amateur imitators. All of these new voices are used as the target voice, the voice the listeners are to identify, in one of the four experiments reported here.

The line-ups used in the earlier Schlichting & Sullivan study were constructed from a range of 10 voices. The familiar voice was a well-known Swedish politician (Carl Bildt, the former *statsminister* of Sweden), a professional imitation of Carl Bildt by Göran Gabrielsson, the natural voice of the imitator, three amateur imitations of Carl Bildt and four natural voices of the amateur imitators. (One of the four volunteer imitators failed to read the entire text as a single speech so only three amateur imitations were used in the experiments.)

Subjects: listeners

The listeners were randomly selected native Swedish speakers between the ages of 19 and 48 with no known hearing damage. They were recruited by advertising and were not paid or awarded course credits. For each test there were 10 female and 5 male listeners. They all indicated that they were familiar with Carl Bildt's voice and not familiar with any of the speakers. This was verified after the experiment had been conducted.

The Material

The phrases presented in the actual line-ups were identical to those used in the earlier studies and were taken from a debate on whether the Baltic States (Lithuania, Latvia and Estonia) should be allowed to join the European Union or not. The phrases were: *och därfor tycker jag att det är så underligt* (and therefore I find it so strange) and *att där vill miljöpartiet bromsa* (that, there the Green Party wants to 'drag its heels'). The recordings of Carl Bildt were received directly from Göran Gabrielsson. All the other recordings were made in the phonetics laboratory's anechoic chamber in the Department of Linguistics, Umeå University. A Panasonic SV-3700 DAT recorder was used with a Neumann U87 microphone and a wind shield. All the recordings were of high quality and of a comparable standard.

The speakers, who attempted to imitate Carl Bildt were given a recording of the passage, a transcript and had 6–8 days in which to perfect their imitation. The same applied to the material used to familiarise the listeners with the voice they were to identify in the line-ups. The text for the

familiarisation of the listeners is taken from a speech Carl Bildt held as the Swedish *Statsminister* in 1993 to the *Riksdag*, the Swedish Parliament. A transcription of both the passage from which the phrases used in the line-ups are taken and the passage used to familiarise the listeners with the voice they were to recognize is to be found in Schlichting & Sullivan (forthcoming).

The construction of the line-ups

The line-ups were six voices in length. Each voice was separated by a less than one second pause (Roebuck & Wilding, 1993): the pause was around 80msec in length. Line-ups were constructed with and without Carl Bildt, with and without the amateur imitations of Carl Bildt to be used as the voice to be recognized, and with and without the natural voices of the amateur imitators, whose imitations were used as the target voices. Foils randomly selected from the set of stimuli clipped from the amateur imitations not to be used as the voice to be recognized and their natural voice recordings were used to make all the line-ups six voices in length. Eight different line-up compositions were, thus, created for each of two chosen stimuli making a total of sixteen different line-ups. The order of voice presentation within the line-up was random.

THE RECOGNITION TASKS

Two different recognition tasks were conducted. The same composition of the set of sixteen line-ups were presented in precisely the same randomised order to the listeners. The difference between the two experiments was the identity of the amateur imitator of Carl Bildt who was used to replace the professional imitation of Carl Bildt and the voice of Göran Gabrielsson in the line-ups. In the first experiment reported here the target voice was of a 22 year old male from Svängsta and in the second a 32 year old male from Gothenburg.

The listeners were twice played a one minute recording of the imitation of Carl Bildt to be recognized. Once before the training block of four line-ups and again before the experimental block of sixteen different line-ups. The passage lasted a different length of time for the two imitators and as in Schlichting & Sullivan's earlier work the passage was repeated until exactly one minute of speech was presented to the listeners. The subjects were informed about the judicial use of the voice line-up technique and were asked to identify the voice they would hear at the beginning of experiment in the voice line-ups there were going to hear. The instructions were identical to the earlier study and can be found in Schlichting & Sullivan (forthcoming). The subjects were asked if they had any questions or required anything clarified between the training and experimental blocks.

RESULTS AND DISCUSSION

The listeners' responses are presented in Tables1–8 in the form of confusion matrices. The tables are to be read as the voice to be identified in the left hand column (the y–axis) and the voice identified by the listeners along the top (the x–axis). In all the tables the boxes representing correct identifications are shaded. The following abbreviations are used in the tables: **C**arl **B**ildt (CB); **P**rofessional **I**mitation (PI); **A**mateur **I**mitation (AI), here AI1 and AI2 refer to the two different amateur imitators; and **N**atural **V**oice of the **I**mitator (NVI). The results relating to the Professional Imitation are taken from Schlichting & Sullivan's earlier work.

In the line-ups containing Carl Bildt's real voice (Tables 1–4) there is a significant difference in the responses of the listeners when identifying the professional as opposed to the amateur imitations. In Schlichting & Sullivan (forthcoming) it is proposed that after having heard the professional imitation of Carl Bildt as the voice to be identified the listeners are convinced that they are to recognize Carl Bildt and respond accordingly. Equally when both Carl Bildt's real voice and the professional imitation are present in the same line-up the listeners are able to discriminate between them and choose the best, ie. Carl Bildt's real voice.

In the case of both of the amateur imitators the listeners do not believe that the voice to be identified is Carl Bildt and respond accordingly. In the case of AI1, Carl Bildt is only selected once and then in the line-up not containing the imitation or AI1's natural voice. A line-up which contains the least comparative information for the listener. In the case of AI2 there is greater confusion between Carl

Bildt, the imitation and AI2's natural voice. This could be due to the Gothenburg accent having a greater degree of similarity to Carl Bildt's accent than the Svängsta accent. This explanation appears to be supported by the spread of the responses across Carl Bildt, the imitation and AI2's natural voice. and, thus, seems more plausible than claiming that AI2's imitation is better than AI1's. This similarity between Carl Bildt, the imitation and AI2's natural voice is also demonstrated in the increase in the selection of both Carl Bildt and the imitation when AI2's natural voice is absent from the line-up. Interestingly the presence or absence of PI's natural voice in the same conditions had no significant effect. This difference in behaviour strengthens Schlichting & Sullivan's (forthcoming) hypothesis that the listeners recognizing the professional imitation decided that the voice they had heard, and had been asked to identify in the line-ups, was Carl Bildt and reacted accordingly. Whereas the listeners in the study reported here who were asked to recognize the amateur imitations were not convinced that the voice they had heard was Carl Bildt and thus they attempted to recognize some quality in the voice they had been asked to identify in the line-ups.

Table 3 and Table 4's confusion matrices represent the forensically plausible line-ups containing Carl Bildt's real voice. It is unlikely that a criminal who has imitated someone will continue to imitate that person in a voice line-up. Although in Table 3 the shaded area is 'Not present' — the listeners were asked to identify the imitation in the line-ups — the forensically 'correct' response is NVI, the natural voice of the imitator. In the case of PI, where the listeners, it is posited, are trying to identify Carl Bildt, they succeed in this and produce judically fatal errors 100% of the time when PI's natural voice is present and 97% of the time when it is absent — an negligible difference. The situation with the amateur imitators differs dramatically. In the case of AI1 50% identify AI1's natural voice and no-one identifies Carl Bildt, when the AI1's natural voice is present. A 50% miss-identification rate is a high judicially incorrect error-rate. However, when AI1's voice is absent from the line-up the error-rate drops to 7%; one of which, however, is an identification of Carl Bildt. The situation regarding AI2 is more complex. AI2 has been more successful in convincing the listeners that he is not present in the line-ups containing his natural voice. AI2's natural voice is only identified 13% of the time, which is comparable with those who identify him as Carl Bildt or one or other of the foils. The judically fatal error rate is 87%. When AI2's natural voice is removed from the line-up leaving only Carl Bildt's voice and foils the percentage choosing 'not present' remains unaltered at 60% which means a judically fatal error-rate of 40%. There is an increase in the selection of Carl Bildt and foils. It does not appear that AI2 has convinced the listeners that they should be recognizing Carl Bildt in the line-ups, yet he has successfully disguised his natural voice.

	CB	Imitation	NVI	Foils	Not present	No. incorrect
PI	25	5				25
	83%	17%				83%
AI1		27	2	1		3
		90%	7%	3%		10%
AI2	2	14	8	3	3	16
	7%	47%	27%	10%	10%	54%

Table 1. The confusion matrix when the line-up contains the voice of Carl Bildt (CB), the target imitation of Carl Bildt (Imitation), the natural voice of the imitator (NVI) and the foils.

	CB	Imitation	NVI	Foils	Not present	No. incorrect
PI	25	4	—		1	26
	83%	13%			3%	86%
AI1		29	—		1	1
		97%			3%	3%
AI2	6	21	—	3		9
	20%	70%		10%		30%

Table 2. The confusion matrix when the line-up contains the voice of Carl Bildt (CB), the target imitation of Carl Bildt (Imitation) and the foils.

The confusion matrices presented in Tables 5–8 are the result of line-ups which do not contain the voice of Carl Bildt. In none of these line-ups is PI's natural voice identified when present in the line-up. PI has been totally successful in shifting recognition from his natural voice. In the line-ups represented in Tables 5 and 7 he was so successful that the judically fatal error-rate is 100%. In the case of the amateur imitators they succeeded, in the line-up represented by Table 5, to have their imitations identified 87% and 73% (AI1 and AI2, respectively) of the time and create judically fatal error-rates of 87% and 93%, respectively. However, when the imitation was removed from the line-up, as presented in Table 7, AI1s natural voice was identified 60% of the time and AI2s 50%. Thus, they have, approximately, a 1-in-2 chance of being correctly identified. They have not been totally successful in shifting recognition from their natural voices, but they have managed to create recognition confusion.

	CB	Imitation	NVI	Foils	Not present	No. incorrect
PI	27	—			3	27
	90%				10%	90%
AI1		—	15	1	14	16
			50%	3%	47%	53%
AI2	5	—	4	3	18	12
	16%		13%	10%	60%	40%

Table 3. The confusion matrix when the line-up contains the voice of Carl Bildt (CB), the target imitation of Carl Bildt (Imitation), the natural voice of the imitator (NVI) and the foils.

	CB	Imitation	NVI	Foils	Not present	No. incorrect
PI	28	—	—	1	1	29
	93%			3%	3%	97%
AI1	1	—	—	1	28	2
	3%			3%	93%	7%
AI2	7	—	—	5	18	12
	23%			17%	60%	40%

Table 4. The confusion matrix when the line-up contains the voice of Carl Bildt (CB) and the foils.

	CB	Imitation	NVI	Foils	Not present	No. incorrect
PI	—	24		1	5	6
		80%		3%	17%	20%
AI1	—	26	4			4
		87%	13%			13%
AI2	—	22	2	3	3	8
		73%	7%	10%	10%	27%

Table 5. The confusion matrix when the line-up contains the target imitation of Carl Bildt (Imitation), the natural voice of the imitator (NVI) and the foils.

	CB	Imitation	NVI	Foils	Not present	No. incorrect
PI	—	13	—	1	16	17
		43%		3%	53%	57%
AI1	—	29	—		1	1
		97%			3%	3%
Ai2	—	21	—	6	3	9
		70%		20%	10%	30%

Table 6. The confusion matrix when the line-up contains the target imitation of Carl Bildt (Imitation) and the foils.

	CB	Imitation	NVI	Foils	Not present	No. incorrect
PI	—	—		3	27	3
				10%	90%	10%
AI1	—	—	18	3	9	21
			60%	10%	30%	70%
AI2	—	—	15	3	12	18
			50%	10%	40%	60%

Table 7. The confusion matrix when the line-up contains the natural voice of the imitator (NVI) and the foils.

	CB	Imitation	NVI	Foils	Not present	No. incorrect
PI	—	—	—	1	29	1
				3%	97%	3%
AI1	—	—	—	7	23	7
				23%	77%	23%
AI2	—	—	—	4	26	4
				13%	87%	13%

Table 8. The confusion matrix when the line-up contains only the foils.

CONCLUSION

The work presented in this paper extends Schlichting & Sullivan's earlier work on the impact imitation can have on the realiablity of the voice line-up. The experiments using two amateur imitiations have been discussed in relation to the results obtained with identical line-up conditions using a high-quality professional imitation. In the case of the professional imitator the listeners are convinced that the recording they hear before the line-ups is Carl Bildt and that it is Carl Bildt that they are to recognize. In the case of the amateur imitations they have not managed to convince the listeners that they are Carl Bildt, but in trying to do so they have managed to disguise their voice to such an extent that for the most judically plausible line-up, shown in Table 7, they only have a 1-in-2 chance of being recognized. (This is the most judically plausible line-up since the imitation will never form part of a real line-up and as they are not mistaken for Carl Bildt, his voice would not form part of the line-up either.) It is, therefore, clear that a high-quality professional imitation of someone else's voice can lead to the identification of the wrong person in a voice line-up. This voice is the one the criminal wishes to be identified. An amateur imitation of another person fails to have the person the criminal wishes to have identified identified, however it leads to a degree of speaker disguise which protects their own identity in a voice line-up using lay listeners.

REFERENCES

Künzel, H.J. (1987) *Sprechererkennung* (Kriminalistik: Heidelberg).

Roebuck, R. & Wilding, J. (1993) "Effects of vowel variety and sample length on identification of a speaker in a line-up", *Applied Cognitive Psychology* 7, 475-481.

Rose, P. & Duncan, S. (1995) "Naïve auditory identification and discrimination of similar voices by familiar listeners", Forensic Linguistics: The International Journal of Speech , Language and the Law 2, 1-17.

Schlichting, F. (1996) *Die imitierte Stimme - Ein Problem für 'voice line-ups'?* C-uppsats in Phonetics, Phonetics Section, Department of Linguistics, Umeå University, Sweden.

Schlichting, F. & Sullivan, K.P.H. (forthcoming) "The imitated voice - a problem for voice line-ups", Forensic Linguistics: The International Journal of Speech, Language and the Law.