

Who knows Carl Bildt? — And what if you don't?

*Elisabeth Zetterholm^{1, 2}, Kirk P H Sullivan^{2, 3}, James Green⁴,
Erik Eriksson², Jan van Doorn⁵ and Peter E. Czigler^{6†}*

¹ Department of Linguistics and Phonetics, Lund University, Sweden

² Department of Philosophy and Linguistics, Umeå University, Sweden

³ Department of Computing Science, Umeå University, Sweden

⁴ Department of Psychology, University of Otago, New Zealand

⁵ Department of Clinical Science, Umeå University, Sweden

⁶ Department of Caring Sciences, Örebro University, Sweden

†the authors are listed in reverse alphabetical order

elisabeth.zetterholm@ling.umu.se, kirk@ling.umu.se,
green@psy.otago.ac.nz, erik.eriksson@ling.umu.se,
jan.vandoorn@logopedi.umu.se, peter.czigler@ivo.oru.se

Abstract

One problem with using speaker identification by witnesses in legal settings is that high quality imitations can result in speaker misidentification. A recent series of experiments has looked at listener acceptance of an imitation of a well known Swedish politician. Results showed that listener expectation of the topic of an imitated passage impacts on the acceptance or rejection of the imitation. The strength of that impact varied according to various listener characteristics, including age of listener. It is likely that age reflected the degree of familiarity with the voice that was being imitated. The present study has reanalyzed the data from Swedish listeners in the previous studies to look at performance according to self reports of whether the listeners were familiar with the politician. Results showed that the acceptance of the imitation by those listeners who reported knowing the politician was more influenced by the topic of the imitated passage than by those who reported not knowing him. Implications of this finding in regard to listeners' choice of alternate voices in the line up are discussed.

1. Introduction

Identification using voice recognition by witnesses can be used as evidence in legal proceedings. However, voice imitation is a threat to the validity of such evidence. A study by Schlichting and Sullivan [1] found that a good imitation of a well known Swedish politician produced a high rate of speaker misidentification [1]. In that study the voice being imitated was the voice of Carl Bildt, a former Prime Minister of Sweden, and the familiarisation passage was a political speech made by Carl Bildt in 1993. It was suggested that the high rate of misidentification may have been due in part to the political topic used in the imitation, increasing the listeners' willingness to accept the imitation as the real voice of Carl Bildt.

Two further studies ([2], [3]), were then designed to test the impact of topic of an imitation on its acceptance by listeners. They found that listener expectation relating to the topic of the imitated passage influenced the acceptance or rejection of an imitated voice. In those studies two imitations of Carl Bildt by the same imitator were used: one with, and one without political content.

The results of both studies showed that listeners were more likely to accept a voice imitation as the actual voice if the topic of the speech passage contained information that was consistent with their expectations. The tendency was stronger for the Swedish listeners in [2] than in [3]. That difference was possibly because listeners in [2] were all young adults from the north of Sweden who were old enough to be familiar with the politician, but who were less familiar with the subtleties of his Swedish accent, while the Swedish listeners in [3] were older adults from the south of Sweden.

To investigate this possibility, yet another study was conducted, replicating [2] and [3], but using listener groups (teenagers and adults) from northern, central and southern Sweden [4]. Results found that region had little impact on the strength of the difference between performance on a political imitation vs non-political imitation. However, the age of the listener groups (adult vs teenage) for the central and southern groups was statistically significant. It is conjectured that this age factor was reflecting the likely familiarity with the politician, as the teenage group would have been very young children when Carl Bildt was Prime Minister of Sweden. This conjecture can be further explored using the existing data from previous studies because all listeners in all studies were asked whether they were familiar with Carl Bildt after they had completed their listening experiments (see method in [2], [3], [4]). Hence the aim of this study was to investigate whether self reported knowledge of the imitated voice influenced acceptance of imitations that were political and non political. Data from all Swedish listeners in the previous studies were reanalyzed, with listeners grouped according to whether or not they reported that they were familiar with Carl Bildt.

2. Method

The data analyzed in this paper is taken from the studies reported in [2], [3], [4] and [5]. All of these studies used the same experimental method. This method is summarized below and the characteristics of the listener groups (as defined for the current investigation) are described in detail.

2.1. The voices

The voices consisted of a set of seven recordings of the same text of a political speech, and one recording of a non-political topic (how to bake a cake). The political recordings were an original by Carl Bildt (PS-Bildt), a professional imitation of the voice of Carl Bildt (AM1), the natural voice of the professional imitator (PS-AM) and four other male voices, referred to hereafter as ‘foils’. The non-political recording was a free voice imitation of Carl Bildt explaining how to bake a cake (AMK). Two of the recordings (AM1 and AMK) were used as familiarization passages, and the remaining six were used as the basis for the voice line-up. Detailed information about the voices used in the experiment, along with the findings of the acoustic and auditory analysis of the recordings, can be found in [2] and [5].

2.2. The listeners

The listeners were all randomly selected and no listener reported any hearing damage. The listeners were native Swedish speakers from Umeå, Örebro and Lund. The listeners from each of these towns were divided into two approximately equal sub-groups.

After the perception experiments, all the participants reported whether they were Familiar with the voice of Carl Bildt (FCB) or not, and if they were familiar with any of the other voices used in the experiment or not. The descriptive statistics of the listener groups is reported in Table 1.

Table 1: The descriptive statistics of the listener groups. Missing gender data points are not explicitly given. Exp 1 heard the AM1 training passage; Exp2 heard the AMK training passage

Exp	No	No. Male	No. Female	Mean Age	SD Age
1	130	51	78	25.30	12.49
2	124	52	68	26.77	13.13

2.3. Familiarization

Two experiments were constructed; each began with a familiarization voice. The experiments differed only in the familiarization passage. Experiment 1 used the political passage imitation (AM1), while Experiment 2 used the “how to bake a cake” passage imitation (AMK).

2.4. The line-up

For both experimental set-ups, the line-up was constructed from the six recordings of the political passage that were not used for the familiarisation task ie recordings by PS-Bildt, PS-AM, and the four other male voices. Three separate segments were spliced out from each of the six recordings. Each segment was repeated three times in the line-up, giving a total of 54 speech stimuli in the line-up (3 repetitions x 3 speech samples x 6 speakers). The line-up voices thus contained PS-Bildt, PS-AM and foils as the test voices.

2.5. Recognition tasks

The two experiments were conducted separately. Both sub-groups were first familiarized with the voice of the target

speaker they were to identify, and were told that they would be asked to recognize it later. Then they listened to a CD containing the 54 speech stimuli, presented in a randomized order. The listeners were instructed to respond ‘Yes’ on a response sheet whenever they recognized the voice from the familiarization recording, and ‘No’ when they did not.

3. Data analysis

The Yes-No data were analyzed from the studies [2], [3], [4] and [5] using techniques from [2], [3], [4], [5] and [6]. The data were categorized as hits (‘Yes’ response to target voice stimulus), misses (‘No’ response to target voice stimulus), false alarms (‘Yes’ response to non-target voice stimulus) and correct rejections (‘No’ response to non-target stimulus) [7].

By taking the number of hits (H) and false alarms (FA), together with the total number of target stimuli and non-target stimuli presented, it is possible to calculate the listeners’ discrimination sensitivity, d' , as follows,

$$d' = z(H) - z(FA) \quad (1)$$

where $z()$ represents the transformation of a proportion to a z-score, H represents the proportion of target trials where the listeners scored a ‘hit’ and FA represents the proportion of non-target trials where the listeners scored a ‘false alarm’. A log-linear correction was applied in the calculation of H and FA as some listener’s detection rates were perfect ($z(H)=\infty$) [8].

A measure of response bias, c , was also calculated to check whether listeners had a tendency to answer ‘yes’ in preference to answer ‘no’, or vice-versa. A positive c value indicates a preference to answer ‘no’; here the false alarm rate is lower than the miss rate. A negative c value indicates a preference to answer ‘yes’; here the false alarm rate is greater than the miss rate: c was calculated as follows:

$$c = [0.5[z(H) + z(FA)]] \quad (2)$$

4. Results

Following [2], [3], [4] and [5], the responses were counted first with the imitator’s (AM’s) voice as the target voice, (that is identification of the natural voice of the imitator being scored as a hit), and then with the voice of Carl Bildt (CB), the voice imitated, as the target voice. The first set of scores shows how well, or badly, the imitator succeeded in convincing the listeners that he was someone else by imitating the voice of Carl Bildt. The second set of scores show how well he convinced the listeners that he was Carl Bildt. Following, [4] and in contrast to [2], [3] and [5] the hits and false alarms were not pooled and using SPSS d' and c -values were calculated for the individual listeners, from which mean d' and c values were computed.

Each sub-group was divided into two further subgroups based upon self-reported knowledge of Carl Bildt. Table 2 reports the descriptive statistics for the resultant subgroups.

The mean d' values for the Swedish listeners divided into Familiar with Carl Bildt’s voice (FCB+) and those Not Familiar with Carl Bildt’s voice (FCB-) are presented in Figure 1.

Table 2: The descriptive statistics of the listener groups. Missing self-reported knowledge of Carl Bildt data points are not explicitly given. Exp 1 heard the AM1 training passage; Exp2 heard the AMK training passage. FCB=Familiar with Carl Bildt.

FCB	Exp	No	No. Male	No. Female	Mean Age	SD Age
Yes	1	104	40	64	26.50	12.79
	2	99	50	46	28.29	13.11
No	1	24	11	13	20.29	10.11
	2	22	1	21	18.14	6.33

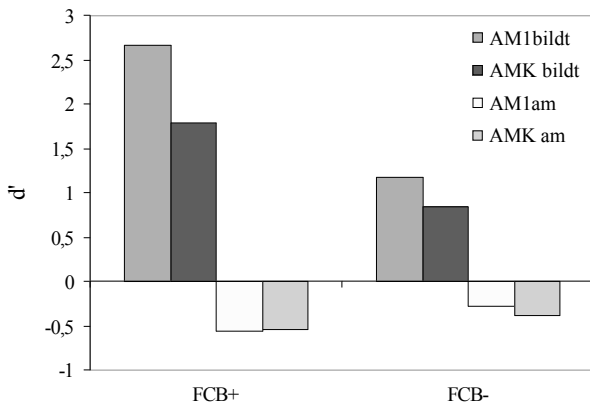


Figure 1: The mean d-prime values for the FCB+ and FCB- listeners. For AM1bildt and AMKbildt Carl Bildt was scored as the target voice. For AM1am and AMKam, AM was scored as the target voice. The prefix AM1 = experiment 1; the prefix AMK = experiment 2

A between-subjects ANOVA (Topic (AM1 or AMK) X Reported CB knowledge (FCB+ or FCB-) was conducted for d' and c values for both CB and AM as target voices. For CB as target, differences in mean d' for both Topic ($F=8.7$, $p<.005$) and FCB ($F=35.1$, $p<.001$) were significant. For AM as target, differences in mean d' reached significance for FCB only ($F=8.5$, $p<.005$). For bias (c) values, there was significant difference (at $p=0.01$) for FCB when results were scored with CB as target. No other bias means were significantly different. There were no significant interaction effects.

Planned post-hoc comparisons on d' showed that for CB as target voice the effect of Topic was significant for the FCB+ group (AM1 Mean $d'=2.66$, AMK Mean $d'=1.79$; $F=22.8$, $p<.001$) but not in the FCB- group. Also, for the group who listened to the political familiarization, the FCB+ listeners performed significantly better than FCB- listeners with both CB (FCB+ Mean $d'=2.66$, FCB- Mean $d'=1.18$; $F=34.8$, $p<.001$), and AM (FCB+ Mean $d'=-0.57$, FCB- Mean $d'=-0.28$; $F=8.2$, $p<.01$) as targets. For listeners of cooking familiarization, FCB+ listeners performed better than FCB- listeners for CB as target only (FCB+ Mean $d'=1.79$, FCB- Mean $d'=0.85$; $F=8.2$, $p<.01$).

The mean c values for the Swedish listeners divided into Familiar with Carl Bildt's voice (FCB+) and those not familiar with Carl Bildt's voice (FCB-) are presented in Figure 2.

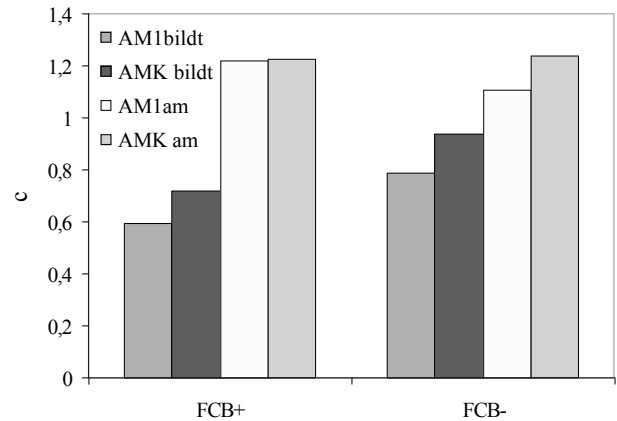


Figure 2: The mean c values for the FCB+ and FCB- listeners. For AM1bildt and AMKbildt Carl Bildt was scored as the target voice. For AM1am and AMKam, AM was scored as the target voice. The prefix AM1 = experiment 1; the prefix AMK = experiment 2

No significant differences were found for c -values in the between-subjects ANOVA (Topic (AM1 or AMK) X Reported CB knowledge (FCB+ or FCB-) for Topic or self-reported CB knowledge.

5. Discussion

The group who reported knowing Carl Bildt and who heard the political familiarization selected CB imitation significantly more often than those who heard the cooking passage. This did not occur for the group who reported not knowing Carl Bildt. The FCB- group did not have any expectation of the sort of topic about which the voice they were asked to remember should talk.

For all listeners who heard the political passage, the FCB+ group performed better (AM1 Mean =2.7 vs. AMK Mean =1.7) than the FCB- group with CB as the target. This could be interpreted by saying that if you are familiar with a voice you will recognize it better (in a line-up). This was also true for the cooking passage listeners, i.e. FCB+ did better than FCB- (at lower d' values) which strengthens the above statement.

This finding has forensic linguistic applications. It shows that it is of importance that the investigators, defense and prosecution lawyers consider a witnesses' degree of topic expectation, that is was the topic the person talked about one that was expected, as well as a witness' familiarity with an individual's voice.

The very small negative d' -values for the recognition AM for the listeners reporting familiarization with Carl Bildt make it difficult to make any claims relating to the found significant effect. The significant effect needs to be confirmed with a reduplication study before its implications can be addressed.

In the light of the finding, that the acceptance of the imitation in the FCB+ was more influenced by the topic of the imitated passage than in the FCB- group, an examination of the listeners' choice of alternate voices was conducted. Differences and similarities were found. Figure 3 shows the distribution of Yes selections for the AM1 and AMK line-ups with no consideration of FCB. Figure 4 shows the Yes selections for the FCB+ listeners and Figure 5 for the FCB- listeners.

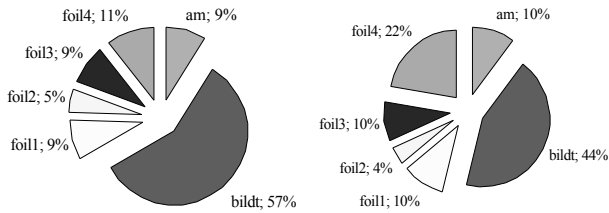


Figure 3: The distribution of selections for each line-up, regardless of self-reported FCB. The chart to the left shows the distribution for the listeners who heard the training passage AM1 (experiment 1); the chart to the right shows the distribution for all listeners who heard the training passage AM1 (experiment 1).

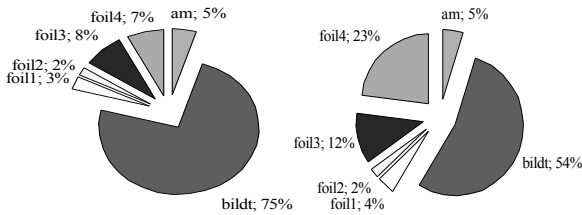


Figure 4: The distribution of selections for each line-up for self-reported FCB+. The chart to the left shows the distribution for the listeners who heard the training passage AM1 (experiment 1); the chart to the right shows the distribution for the listeners who heard the training passage AM1 (experiment 1)

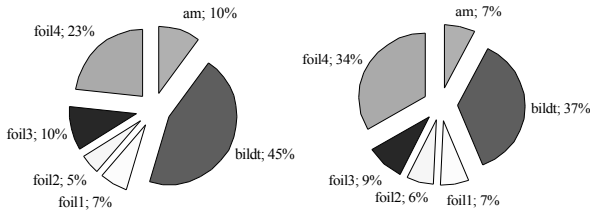


Figure 5: The distribution of selections for each line-up for self-reported FCB-. The chart to the left shows the distribution for the listeners who heard the training passage AM1 (experiment 1); the chart to the right shows the distribution for the listeners who heard the training passage AM1 (experiment 1)

For all AMI listeners (Figure 3) the distribution of non-CB Yes selections is approximately evenly distributed across am and the four foils. For the AMK listeners, foil4 is selected more frequently than am or any other foil. With knowledge of Carl Bildt (Figure 4), the AMI listener non-CB-Yes-responses are no longer evenly distributed with foil3 and foil4 being marginally selected more frequently than am or the other two foils. For the AMK listeners the non-CB-Yes-responses are most common for foil4. An almost identical Yes-response frequency is found for the self-reported FCB- groups having heard the AMI passage (Figure 5). The percentage of Yes-selections of Foil4's was largest for the AMK FCB- listeners.

These figures show the relative distribution of the Yes-responses. From these distributions, it can be seen that one or more features of foil4 form part of the feature set that the listeners are attending to when making their Yes-No selections. Further research will include close acoustic examination of this foil. The figures also show the danger

of basing a legal argument solely on positive responses by a witness; the difference in Yes-responses for CB in Figure 5 looks large, yet the d' value for these two listener groups was not found to be statistically significantly different. The complete response pattern by a voice witness needs to form the basis of voice based evidence.

6. Conclusions

The results of this study show that knowledge of a speaker, in this case Carl Bildt, does impact upon the selection of that voice in a Yes-No two-way alternative forced choice test. The listeners with a knowledge of Carl Bildt's voice had significantly higher d' values, that those who self-reported no knowledge of Carl Bildt's voice. Further the topic of the training passage resulted in significantly different responses for those who reported familiarity with Carl Bildt's voice, yet not for those who reported no familiarity with Carl Bildt's voice. The danger of not considering the entire response pattern of a voice witness has also been demonstrated.

7. Acknowledgments

This research was supported by The Bank for Sweden Tercentenary Foundation through their funding of the project "Imitated voices: A research project with applications for security and the law" Dnr K2002-1121:1, Dnr K2002-1121:2.

We thank the Umeå University cognitive science students who helped us with the data collection and all the participants who took part in the study.

8. References

- [1] Schlichting, F. and Sullivan, K.P.H., "The imitated voice — a problem for voice line-ups?", *Forensic Linguistics, Vol. 4, 1997, pp.148–165*.
- [2] Zetterholm, E., Sullivan, K.P.H. and van Doorn, J., "The impact of semantic expectation on the acceptance of a voice imitation", *SST2002 Proc.*, Melbourne, Australia, pp. 379–384, 2002.
- [3] Sullivan, K. P. H., Zetterholm, E., van Doorn, J., Green, J., Kügler, F. and Eriksson, E., "The effect of removing semantic information upon the impact of a voice imitation", *SST2002 Proc.*, Melbourne, Australia, pp. 291–296, 2002.
- [4] Zetterholm, E., Sullivan, K. P. H., Green, J., Eriksson, E. and Czigler, P. E., "Imitation, expectation and acceptance: The role of age and first language in a nordic setting", *ICPhS 2003 Proc.*, In press.
- [5] Zetterholm, E., *Voice imitation. A phonetic study of perceptual illusions and acoustic success*, Travaux del l'institut de linguistique de Lund, Lund University Press, Lund, 2003.
- [6] Schiller, N.O., Köster, O. and Duckworth, M., "The effect of removing linguistic information upon identifying speakers of a foreign language", *Forensic Linguistics, Vol. 4, 1997, pp. 1–17*.
- [7] Green, D.M. and Sweets, J.A., *Signal Detection Theory and Psychophysics*, New York: Wiley, 1966.
- [8] Hautus, M. J. "Corrections for extreme proportions and their biasing effects on estimated values of d' ", *Behaviour Research Methods, Instruments and Computers, Vol. 27, 1995, pp.46–51*.