PERCEPTUAL RECOGNITION OF FAMILIAR VOICES USING FALSETTO AS A TYPE OF VOICE DISGUISE

Isolde Wagner and Olaf Köster
Bundeskriminalamt, Speaker Identification Department, Wiesbaden, Germany

ABSTRACT

This paper discusses the use of falsetto phonation as a form of voice disguise. A perceptual recognition experiment has been carried out. In a naming task it was tested whether phonetically untrained listeners are able to identify familiar speakers when falsetto disguise is involved. It was found that overall performance of identification is very high for the normal speaking condition but significantly lowered for the condition of falsetto phonation. Recognition rates are reduced from 97% in the case of undisguised voices to merely 4% for falsetto as a disguise technique. The results of this experiment are in accordance with the findings of earlier studies on the recognition of different types of disguised voices and confirm the importance of phonetically trained experts in the analysis of disputed speech material.

1. INTRODUCTION

Voice disguise is a problem frequently met in the field of forensic phonetics. It potentially occurs in all criminal cases in which an anonymous caller is concious of either committing a crime (e.g. blackmail) or giving information concerning a crime (e.g. in cases of kidnapping and homicide) and is aware of the fact that his/her voice may be used for identification. Following an earlier study [7] where subjects were asked to disguise their voices and the majority of disguises were found to be an alteration of phonation, a recent research study based upon German forensic casework [11] has shown that one of the most popular means of voice disguise is the use of falsetto voice quality. For example, in a spectacular criminal case the blackmailer in nearly 40 telephone calls constantly spoke with falsetto phonation. Contrary to normal practice in Germany the material was not released to the public for recognition purposes because the fundamental question arose as to whether voice recognition by familiar persons is at all possible when falsetto disguise is involved.

Compared to modal voice, falsetto is discribed to be based on completely different laryngeal vibratory patterns. The vocalis muscles along the glottal edge of each vocal fold remain relaxed, but the mass of each vocal fold is made stiff and immobile. The vocal ligaments along the glottal edge of the vocal folds are put under strong tension. This results in the vertical cross-section of the edges of the vocal folds becoming thin. The glottis often remains slightly apart, and the characteristic sub-glottal air pressure is lowered. Only the thin margins of the vocal folds participate in phonatory vibration. The different physiological issues result in different acoustical factors: the first is a considerably higher fundamental frequency, the second is the fact that harmonically-related overtones are widely separated in frequency, and consequently in any given frequency range there will be fewer components in the sound produced than there is in a voice with a lower fundamental frequency. The third is a steeper opening portion of the slope of the laryngeal waveform compared to modal voice [5] (p. 118 ff). Thus, whereas speaking with falsetto phonation the natural vocal behaviour is masked to a high degree and recognition even of quite familiar speakers is supposed to be nearly impossible.

Relatively little research has been done on the recognition of disguised voices [2] (p.196). Some studies report that speaker disguise reduces identification accuracy [1,3]. Although the presence of disguise is generally detected with a high degree of accuracy and reliability [8], Reich and Duke [9] found in a discrimination test that speaker recognition rates are reduced from 92% in the case of undisguised voices to 59-81% for disguised voices. The use of falsetto voice as a disguise technique has not been examined in detail to date.

2. EXPERIMENT

In order to test the recognition ability of phonetically untrained listeners for the condition of falsetto phonation in comparison to natural voice quality, the following experiment has been carried out: in a familiar speaker naming task [4] listeners had to identify familiar speakers from a randomized set of different voices. The voices consisted of both known persons giving the target voices and unknown persons representing the foils (dummy speakers). All persons were recorded with their normal voices and using falsetto as a form of speaker disguise.

2.1. Speech Material

Five male speakers who were colleagues well known to the listeners served as target speakers. They were between 48 and 62 years of age and had no prominent regional accent and no distinctive habits of speaking. Three additional male German speakers who were unknown to the listeners served as foils. They were 38, 50 and 52 years of age and fulfilled the same speaking conditions as the target speakers.

Each of the eight speakers was asked to read the text of a blackmailer's telephone call using both normal and falsetto phonation. The speech samples were about 15 seconds in duration, and thus considererd to be sufficient for auditory recognition purposes [4] (p. 95). To simulate speech samples with the acoustic conditions usually found in forensic recordings, the speech samples were recorded with a DAT-recorder SONY PCM-M1 via telephone transmission. The complete test material consisted of a tape recording of 48 randomized speech samples: 8 speakers x 2 conditions of voice

quality (normal/ falsetto) x 3 repetitions.

2.2. Subjects

The group of listeners consisted of scientific employees at the German Bundeskriminalamt (federal criminal office) working in different sections of the criminal science laboratories. The participants were between 36 and 62 years of age and all of them were phonetically untrained listeners and unexperienced in professional speaker identification. None of them reported any hearing problems.

2.3 Method

The subjects were instructed to listen carefully to the test tape. The speech samples were presented by loud-speakers in a quiet room. The subjects were told to indicate the names of those persons whose voices they had recognized as familiar. It was not mentioned that falsetto was involved in addition to normal phonation. After every voice token a pause of 5 seconds was provided to enable the listeners to make their decisions which they marked on a response sheet. The total experiment lasted approximately 20 minutes.

After the recognition task, a second rating scale response sheet was handed out to the listeners. They were asked to indicate the degree of familiarity to the different target speakers. The rating scale comprised 4 points: (0) I do not know this person/ the voice of this person at all, (1) I know this person/ the voice of this person briefly/ for a short time, (2) I know this person/ the voice of this person well/ for a long time, (3) I know this person/ the voice of this person very well/ for a very long time. Only 20 subjects (19 males and 1 female) who indicated that they knew all of the target speakers with a degree of (2) or (3) were taken into account.

3. RESULTS

In a naming task experiment, listeners can produce two different categories of both correct answers and errors: on the one hand a target speaker can either be correctly identified ("hit") or incorrectly rejected as being unknown ("miss"), on the other hand a dummy speaker can either be correctly rejected as being unknown ("correct rejection") or incorrectly identified as a target speaker ("false alarm"). The proportion of hits and misses is defined as the hit rate (H); the proportion of correct rejections and false alarms is defined as the false alarm rate (F).

3.1. Recognition Performance

Contrasting the two speaking conditions of normal and falsetto voice, a prominent difference in recognition performances was noted. Whereas in the undisguised, normal voice condition 97% of all familiar speakers are correctly identified, only 4% of the same speakers were correctly identified in the falsetto condition.

Figure 1 shows the recognition performance for every target speaker separately. It can be seen that there are differences in the identification of individual targets.

3.2. Discrimination Ability

Hit rate and false alarm rate were used to determine the subjects' sensitivity to the target-dummy-difference, which expresses the listeners'ability to discriminate between a familiar speaker and an unknown person. The overall performance of identification has been calculated by using Signal Detection Theory (SDT) [6]. In SDT, the performance of identification is expressed by the sensitivity measure *d'*. It takes into account both hit rate and false alarm rate, and increases when either H

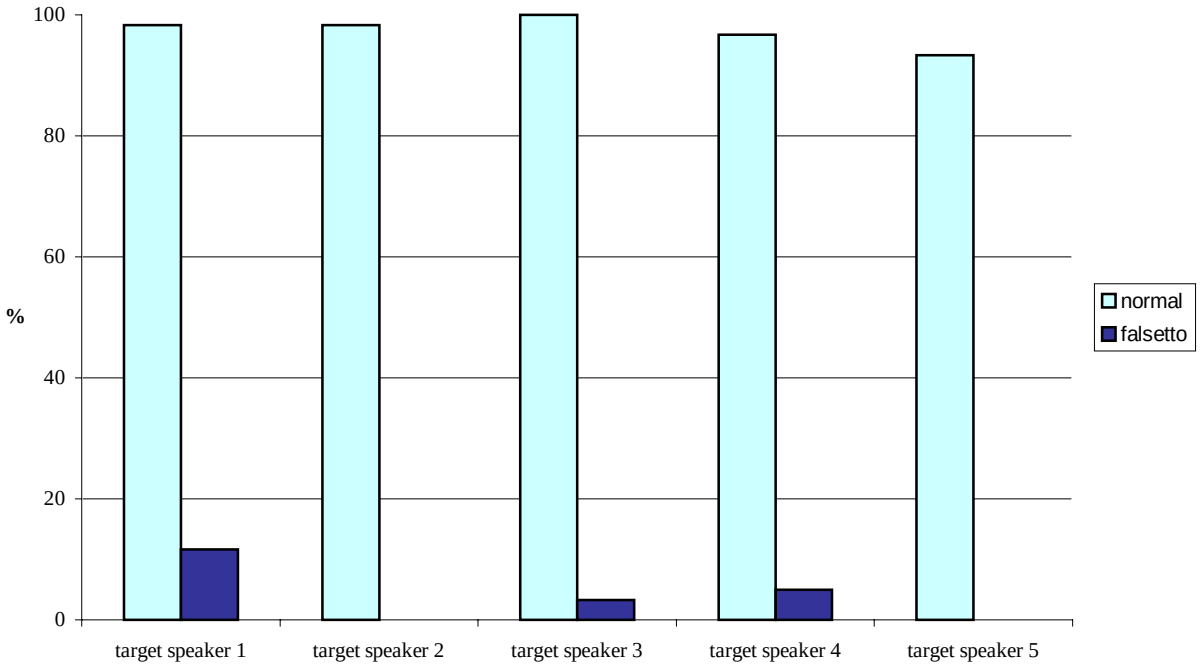

Figure 1: Correctly identified target speakers under normal and falsetto voice condition

increases or F decreases. *d'*-values can range from 0.0 (chance level) to 3.1 (about 100% correct recognition). A more detailed description of this statistic procedure can be found in Schiller, Köster, Duckworth [10].

As can be seen in Table 1, in the normal speaking condition the discrimination ability between target and dummy speakers is very high, whereas in the condition of falsetto discrimination ability between the same speakers decreases to nearly chance level. Further statistical comparison between the two different speaking conditions showed that there is significantly better performance in the recognition of the normal voice quality compared to the condition of falsetto phonation ($p<0.05$).

voice quality	H	F	*d'*
normal	0.97	0.29	2.434
falsetto	0.04	0.01	0.408

Table 1. Hit rates (H), false alarm rates (F) and sensitivity measure *d'* under normal and falsetto voice quality conditions.

Concerning the individual recognition performance, it can be noted that the 20 listeners generally performed in a similar way. In the undisguised, normal voice condition, the lowest score for a subject was 13 hits out of 15 target stimuli. In the falsetto voice condition, the highest score for a subject was 3 hits out of 15 target stimuli.

4. DISCUSSION

Recognition performances under the two conditions were the reverse of one another. Whereas there were only 3% of misses for the normal voice condition, subjects recognized only 4% of the same targets speaking with falsetto phonation.

Furthermore, subjects tend to be less critical when trying to recognize familiar persons speaking with their normal voices. This can be seen from the high score of 29% of false alarms. In these cases the subjects wrongly identified a dummy speaker as either one of the target speakers or as another working colleague who actually did not take part in the experiment at all. The high false alarm rate could be explained by the subjects' wrongly biased expectations. Generally, it could be concluded for the undisguised speaking condition, that people are able to reliably recognize familiar speakers from a 15 second sample even though the acoustic quality of the recording had been degraded by telephone transmission.

The low false alarm rate of only 1% in the cases of falsetto disguised voices may be explained by the fact that in general subjects could hardly recognize familiar speakers or discriminate between unfamiliar and familiar speakers, respectively.

Concerning the recognizability of individual target speakers, as can be seen from Figure 1, all targets were recognized almost equally well in the normal voice condition, but there were more differences in the falsetto voice condition. Target speakers 2 and 5 were not recognized at all. Target speaker 1 was recognized considerably better than target speakers 3 and 4 (7 times out of 60 possible correct identifications, corresponding to a hit rate of 0.12). This cannot be assumed to be the result of a higher degree of familiarity, but probably as a result of the fact that this speaker raised his voice less than the other speakers (Table 2 shows the single values of the mean fundamental frequency for each speaker). It can be concluded that, although even small changes in pitch may have great affects on overall performance of identification, the degree of recognition ability seems to correlate with the extent of changes in the fundamental frequency.

speakers	mean fo (Hz)	standard dev. of mean fo (Hz)	articulation rate (syllables/sec)
targets			
t 1 normal	110	15	4.9
t 1 falsetto	210	33	4.9
t 2 normal	125	17	3.8
t 2 falsetto	287	41	3.6
t 3 normal	118	15	4.6
t 3 falsetto	342	49	4.9
t 4 normal	111	17	4.9
t 4 falsetto	341	42	4.3
t 5 normal	98	17	5.8
t 5 falsetto	249	65	4.3
foils			
f 1 normal	102	18	5.3
f 1 falsetto	352	81	4.1
f 2 normal	128	13	5.3
f 2 falsetto	320	58	4.6
f 3 normal	104	19	3.8
f 3 falsetto	298	50	3.6

Table 2. Mean fundamental frequency, standard deviation and articulation rate of target and dummy speakers under normal and falsetto voice quality conditions.

As can be seen from Table 2, speakers additionally changed their standard deviation of fundamental frequency and their articulation rates when speaking with falsetto phonation. For some speakers, also differences in intonational patterns were observed when changing from modal to falsetto voice. It would appear that when falsetto is used for voice disguise several other vocal features may be affected and thus individual speech features may significantly be masked. This conclusion is based on the very low overall performance of identification of the disguised voices in the perception experiment.

5. CONCLUSION

To conclude, for phonetically untrained listeners it seems very difficult or almost impossible to recognize a familiar speaker when falsetto disguise is adopted. Recognition performance for the same speaker speaking with his habitual (modal) voice quality is highly reliable even though acoustic quality is reduced by telephone transmission. These findings are in general accordance with earlier studies examining different types of disguise techniques [1,3,9].

One inference which might be drawn from the results, is that auditory testimonies of phonetically untrained persons must be judged extremely cautiously and voice recordings should not be released to the public for recognition purposes when falsetto

disguise is involved. However, further research is required in the field of voice disguise.

ACKNOWLEDGMENTS

We thank Herbert Masthoff and Allan Hirson for their helpful comments on the paper.

REFERENCES

[1] Hirson, A., Duckworth, M. 1995. Forensic Implications of Vocal creak as Voice Disguise. *Beiträge zur Phonetik und Linguistik, 64: Studies in Forensic Phonetics*, 67-76.

[2] Hollien, H. 1990. The Acoustics of Crime. New York: Plenum Press.

[3] Hollien, H., Majewski, W., Doherty, E.T. 1982. Perceptual Identification of Voices under Normal, Stress and Disguised Speaking Conditions. *Journal of Phonetics*, 10, 139-148.

[4] Künzel, H. 1990. Phonetische Untersuchungen zur Sprecher-Erkennung durch linguistisch naive Personen. *Zeitschrift für Dialektologie und Linguistik*, Beihefte, Heft 69, Stuttgart: Franz Steiner Verlag.

[5] Laver, J. 1980. The Phonetic Description of Voice Quality. Cambridge: Cambridge University Press.

[6] MacMillan, N.A., Creelman, C.D. 1991. Detection Theory: A User's Guide. Cambridge: Cambridge University Press.

[7] Masthoff, H. 1996. A Report on a Voice Disguise Experiment. *Forensic Linguistics*, 3, 160-167.

[8] Reich, A.R. 1981. Detecting the Presence of Disguise in the Male Voice. *Journal of the Acoustical Society of America*, 69, 1458-1468.

[9] Reich, A.R., Duke, J.E. 1979. Effects of Selected Vocal Disguise upon Speaker Identification by Listening. *Journal of the Acoustical Society of America*, 66, 1023-1028.

[10] Schiller, N., Köster, O., Duckworth, M. 1997. The effect of removing linguistic information upon identifying speakers of a foreign language. *Forensic Linguistics*, 4, 1-17.

[11] Wagner, I. (in preparation). Recent Trends in Voice Disguise. Paper presented to the International Association for Forensic Phonetics' conference 1998, Rijswijk, The Netherlands.