

THE SIGNIFICANCE OF PHONETICS IN VOICE IMITATION

Elisabeth Zetterholm

Dept. of Linguistics and Phonetics, Lund University, Sweden

ABSTRACT: Speech behaviour and the voice show our regional, social and personal identity. Sometimes we imitate other people's speech behaviour in the aim of learning another language or accent or for entertainment. This study indicates that it is possible to imitate other speaker's voice and speech behaviour with success. The result of a perception test indicates that some features, such as voice quality and pitch register, are more important than others for voice identification.

INTRODUCTION

Our voices and our speech behaviour are cues for the listener about the speaker's identity and make it possible for the listener to recognise voices even without seeing the speaker. Some of these features represent the regional and social dialect, but some features are individual. A professional impersonator, who reproduces another speaker's voice and speech behaviour, has to be aware of how to change the vocal tract and get close to the voice of the target speaker. To succeed with the voice imitation he has to figure out important and characteristic features of the target voices. Referring to Laver (1994) imitation, mimicry for example, is a stereotyping process and not an exact copy of the target speaker. In view of that the impersonator may fail in imitating some features if he is successful in imitating other more critical features.

Aims of the present study

To find out how flexible the human voice is and to try to figure out important features for a successful voice imitation, one impersonator and a number of his different voice imitations have been studied. Analysis, both auditory and acoustic, has been done to see how much he changes his own voice. In order to find out individual characteristic features in speech one has to pay attention to both segmental, suprasegmental as well as extralinguistic aspects. There is no comparable analysis with the target voices in this study. A whole utterance as well as one neutral word has been analysed. A perception test has been done in order to get an objective judgement of the voice imitations. 20 participants were asked to grade the voice imitations and comment on their grades. The purpose in this test was not to identify the original voices, just to grade the voice imitations, and the names of the target speakers were given in the test.

MATERIAL

A Swedish professional impersonator has recorded a tape for the Swedish telephone company Telia, containing imitations of some well-known Swedish people. Some of them are politicians and some of the target speakers are well-known TV-hosts or newscasters. 12 voices have been selected for this study and the recordings of each voice vary from 9 to 12 seconds.

The texts were created, by the impersonator, to suit the vocabulary and other features of the target speaker. That should make it easier for the listener to recognise the voices but the main goal is of course entertainment. In all the recordings one neutral word, which is not a typical or even a frequently used word for any of the target speakers, occurs. That is the word 'mobilsvar' (the mobile phone answering machine). As the prejudice seems to be that a successful impersonation depends highly on the choice of vocabulary items, speech style and contextual features, the specific phonetic study of such a neutral word in a number of voice disguises may tell us something about the validity of this prejudice.

CHARACTERISTIC FEATURES

Our identity, group and personal identity, may be communicated by our voices (Pittam 1994). A great deal of a person's speech behaviour is learned. The speakers pick up the regional and social accent and learn to pronounce the sound segments in the same way as other members of the community. There are still

differences between speakers depending on anatomical differences and individual phonetic habits in articulation and speech production, e.g. voice quality, pitch register and speech style.

Swedish dialects

The target voices represent different Swedish dialects, especially from the south, east and west of Sweden. The categorisation of the Swedish dialects refers to Bruce & Gårding's (1978) prosodic accent typology for Swedish dialects. There are obvious differences between the dialects concerning both the intonation pattern and segments, particularly the r-segment. Dialects in south Sweden and some of the dialects in west Sweden use a uvular [ʀ] or [ʁ]. Dialects in the east and north of Sweden use an alveolar trill [r] (Eler 1966). There are individual pronunciations of the r-segment as well, like a fricative [z] or more like a [ə].

Defining 'speech style'

In this study 'speech style' refers to speech tempo, speech rhythm, articulation and continuity. In defining speech rhythm the musical term 'staccato' will be used for some of the voices, in the sense of a rhythmic pattern in a phrase or meaning divided into short phrases, sometimes with a fast speech tempo, separated with pauses. That means that there are more short pauses than necessary for breathing or syntactically motivated. Speech rhythm also correlates with the prosodic pattern of interaction between stressed and unstressed syllables. Some of the target speaker use hedges, hesitation sounds and filled pauses and this is also a part of a speaker's speech style.

RESULTS

In this paper a brief summary of the auditory analysis and some of the results from the acoustic analysis will be presented. For further details see Zetterholm (forthc.) The result of both the auditory and the acoustic analysis refers to the whole utterance as well as the target word 'mobilsvar'. The imitated voices will be referred to only with their initials. GG is the impersonator's own voice, for comparison.

Auditory analysis

An analysis has been done focusing on the dialect, the speech style and intonation pattern as well as voice quality, different characteristic features and pronunciations of some segments.

The impersonator seems to capture the different dialects very well with regard to intonation patterns and different pronunciation of the sound segments, the r-segment as well as vowels (Bruce 1998, Bruce & Gårding 1978). He captures the regional dialects but it is still possible to recognise the individual features of the target speakers from the same area. None of the target speakers have a dialect from the same area as the impersonator, the transition area between east and west of Sweden.

The different speech styles are obvious and sometimes exaggerated in the voice imitations. A fast speech tempo and a slurred articulation, a slower speech tempo with hesitation sounds as well as a staccato-like speech rhythm are characteristic features of some of the target speakers. The impression is that the impersonator has captured the individual speech style of the target speakers. The individual modal voice quality is hard to describe but the impression is that the impersonator is successful in imitating the different voice qualities. Only the most common terms and obvious auditory impression of voice quality, like nasal, breathy, creaky and tense voice quality has been analysed.

The texts in these imitations are created to suit the vocabulary and other features of the target speaker. In using characteristic words and phrases the impersonator strengthens the impression of the imitations. Some of the features are exaggerated and sometimes more like a caricature of the target speaker. Individual features, like hedges, hesitation sounds and filled pauses (Stenström 1999) are clear in the voice imitations. The different pronunciation of the r-segment depends on regional and social differences but there are also individual differences. There are also differences in the pronunciation of the s-segment between the original speakers and this is also evident in the voice imitations.

Acoustic analysis

In the acoustic analysis the mean value of the fundamental frequency in the whole utterance as well as for the target word 'mobilsvar' has been measured. Waveform and spectrogram have been analysed in order to compare the auditory analysis according to voice quality and specific segments. The results of the different voice imitations are compared to the impersonator's own voice. For the acoustic analysis the speech analysis programme Praat has been used.

The average fundamental frequency of the whole utterance (see table 1) as well as the target word 'mobilsvar' (see table 2) shows a great variety between the voice imitations. That indicates that the impersonator is flexible concerning fundamental frequency. The results correspond to the auditory impression, the creaky voices, IW and BÖ for example, are lower and the tense voices, CB and IK, are higher in this investigation according to the acoustic analysis. The duration values depend on the variety of the texts and the speech style in the voice imitations and give a rough idea about duration differences.

Table 1. Duration, F0 mean and std dev. in the whole utterance.

	Duration	F0 mean	F0 std dev
GG	11,45 s	118 Hz	24 Hz
AB	8,86 s	165 Hz	26 Hz
AS	12,13 s	142 Hz	35 Hz
BK	14,00 s	127 Hz	19 Hz
BÖ	11,99 s	105 Hz	28 Hz
CB	21,78 s	215 Hz	41 Hz
CL	10,59 s	123 Hz	31 Hz
HT	11,24 s	139 Hz	30 Hz
HV	15,73 s	119 Hz	36 Hz
IK	12,08 s	255 Hz	40 Hz
IW	10,80 s	97 Hz	16 Hz
KO	18,01 s	151 Hz	52 Hz
OJ	11,33 s	127 Hz	22 Hz

Table 2. Duration, F0 mean and std dev. for the target word 'mobilsvar'.

	Duration	F0 mean	F0 std dev
GG	0,88 s	116 Hz	27 Hz
AB	0,96 s	163 Hz	21 Hz
AS	1,28 s	138 Hz	38 Hz
BK	0,79 s	151 Hz	12 Hz
BÖ	0,81 s	96 Hz	18 Hz
CB	1,28 s	207 Hz	39 Hz
CL	1,00 s	126 Hz	16 Hz
HT	0,96 s	122 Hz	17 Hz
HV	0,92 s	104 Hz	25 Hz
IK	0,94 s	243 Hz	28 Hz
IW	1,31 s	99 Hz	16 Hz
KO	1,04 s	137 Hz	41 Hz
OJ	0,66 s	133 Hz	35 Hz

Different voice qualities can be seen both in the waveform (Zetterholm 1999) and the spectrogram. According to the auditory analysis the creaky voice qualities are observed in the waveforms, as well as in the spectrograms. The auditory impression of breathy and tense voice qualities in some of the imitations is also obvious in the acoustic analysis.

The formant frequencies, F1-F4, of the stressed vowels [i:] and [a:] of the target word 'mobilsvar' have been measured and a comparison made between the imitated voices and the impersonator's own voice. The measurements have been done, every 10 ms, in the Praat programme and the average value for each formant is presented in table 3 and 4. There are individual differences, but it is hard to find some uniform pattern for voices representing the same dialect. The measurement of the vowel length also present individual differences. The individual duration differences of the vowel [a:] are bigger than for the vowel [i:]. One explanation can be that the a-vowels with the longest duration occur when the word 'mobilsvar' is focused at the end of a phrase.

The duration of different segments, the stressed vowels [i:] and [a:] and the r-segment for example, in the target word 'mobilsvar' are clear in the spectrograms as well. There are also differences in the pronunciation of the s-segment in the target word according to the auditory analysis. The lower limit of the frequencies of /s/ are different, but no measurement of the s-segments has been done.

Table 3. The duration and average value of the formant frequencies of the vowel [i:], in Hz.

	Duration	F1	F2	F3	F4
GG	0,135 s	342,79	2142,43	2806,25	3308,04
AB	0,098 s	366,68	1989,27	2947,92	3094,50
AS	0,164 s	284,35	2234,71	2504,13	3536,20
BK	0,112 s	485,89	2018,71	2824,53	3618,76
BÖ	0,105 s	363,73	2105,62	3129,78	3739,21
CB	0,155 s	486,26	1975,48	2922,46	3721,63
CL	0,102 s	347,49	2397,77	3000,72	4525,10
HT	0,152 s	299,17	2174,92	2969,70	3565,12
HV	0,143 s	279,19	2198,45	3214,73	4338,37
IK	0,110 s	282,85	2535,20	2910,28	4360,24
IW	0,176 s	363,29	1863,56	3006,78	3252,52
KO	0,087 s	349,61	2107,60	2965,72	3801,99
OJ	0,099 s	275,00	2304,51	2754,34	3762,19

Table 4. The duration and average value of the formant frequencies of the vowel [ɑ:], in Hz.

	Duration	F1	F2	F3	F4
GG	0,133 s	590,32	886,08	2300,29	3109,00
AB	0,212 s	761,49	1041,11	2561,98	3528,95
AS	0,177 s	515,80	843,73	2790,69	3484,93
BK	0,048 s	677,45	1358,93	2635,08	3488,10
BÖ	0,101 s	628,16	997,59	2704,91	3371,47
CB	0,293 s	758,97	1064,08	2680,88	3472,64
CL	0,083 s	586,78	1027,38	2746,61	4489,79
HT	0,085 s	447,34	835,36	2528,47	3398,83
HV	0,145 s	525,21	830,30	2545,05	3057,86
IK	0,196 s	673,96	1061,96	2485,83	3456,22
IW	0,229 s	808,45	1102,04	2737,86	3477,04
KO	0,143 s	507,23	877,90	2288,07	3452,35
OJ	0,059 s	519,24	784,88	2672,21	3513,59

PERCEPTION TEST

In order to get an objective judgement of the voice imitations a perception test was done. The test was designed on a unix workstation. The participants were asked to grade the voice imitations according to degree of success. The names of the target voices were given since the purpose of the test was not to identify the original voices. The test consisted of two parts. In the first part the participants listened to the target word, 'mobilsvar', and in the second part they listened to the whole utterance with the imitated voices. There were no comparison recordings with the original voices. The listeners had to grade the voice imitations on a scale of 1-5. 1 is not like the original voice at all and 5 is very close to the target speaker. The listener could choose 'unknown' if he/she was not familiar with the target voice. The participants were requested to comment on their judgements. It was possible to listen to the voice imitations in any order in each part of the test.

Results of the perception test

Generally the results show that most of the imitations received a higher grading in the second part of the test (whole utterance) than in the first part (one word), but there is not typically a big difference. The mean value of the judgement for each voice imitation is presented in table 5 and 6, graded according to ranking

order. The number of 'unknown' voices is lower in the second part of the test with the whole utterance, 23 'unknown' compared to 27 'unknown' in the first part of the test, one word.

Table 5. The mean value of the target word 'mobilsvar' in the perception test.

HV	4,58
CB	3,90
CL	3,69
IW	3,61
BÖ	3,45
IK	3,40
KO	3,26
AB	2,83
OJ	2,79
HT	2,78
BK	2,67
AS	2,45

Table 6. The mean value of the whole utterance in the perception test.

HV	4,75
KO	4,21
CL	4,00
CB	3,95
IW	3,78
HT	3,24
IK	3,25
AS	3,00
BÖ	3,00
BK	2,95
OJ	2,80
AB	2,44

Comments from the listeners

According to comments from the participants some of the voice imitations are exaggerated, some more like a caricature. That makes the voice imitation convincing and is in favour in some cases (CB and IW for example). Exaggeration and not exact pronunciation of segments do not affect the overall impression when judging the whole utterance. The voice imitation of HV has the highest mean value in both tests. The impersonator is successful with voice quality, pitch register, intonation and speech style, such as speech tempo and rhythm of the target speaker. According to comments from the listeners the voice imitation is very close to the target voice. The imitations of CB and CL also have high mean values in both tests. The listeners comment the ranking based on the speech style, speech rate and articulation, which is close to the target speakers. The target speaker CB has a characteristic pronunciation of the r-segment and intonation pattern on stressed syllables and that is important too in the voice imitation. The imitation of KO got a higher mean value for the whole utterance compared to the test with one word. According to the comments from the participants the prosody (intonation and phrasing), pitch register, the speech style and the dialect is important in this voice imitation. The text in the imitation of KO is very important and plays a great role for the ranking. Some of the voice imitations at the bottom of the ranking list have a voice quality and pitch register (AB, AS, BK) that do not correspond closely to the target speakers, according to the comments. In the voice imitation of OJ the participants comment that the impersonator fails with the voice quality and the articulation.

CONCLUDING REMARKS

The human voice is flexible and it is possible to change the vocal tract consciously in order to acquire new phonetic behaviour. For voice identification phonetic habits, such as voice quality, intonation pattern and speech style, are important since that is a part of a speaker's individual style. The voice imitations are not an exact copy but the impersonator manages to pick up characteristic features of the target voices and the listeners can recognise the target speakers. Almost all of the target speakers in this analysis have a dialect from the southern part of Sweden. There are obvious differences between these dialects, especially the pronunciation of the r-segment, but even the intonation pattern differs. In this study there are differences between the voices depending on the regional dialect but speakers from the same area show obvious individual differences as well. There are evident individual differences in the voice imitations and it is hard to find any uniform pattern compared to the dialect of the target speakers. The impersonator captures the individual differences very well in his voice imitations. The individual style is a combination of the regional and social identity. The speech style, tempo, rhythm, articulation and continuity, vary between the target speakers in this analysis. In these imitations the impersonator shows

his flexibility and we realise, according to the comments from the participants in the perception test, the importance of speech style in identifying the target voices. Voice quality is a part of our personality but it is hard to describe different voice qualities in normal voices (Zetterholm 1999). There is no useful terminology for different voice qualities in normal voices. According to the results in the perception test voice quality seems to be an important feature for a successful voice imitation. The auditory impression of pitch and voice quality is confirmed in the acoustic analysis. According to the auditory analysis the voices with a creaky voice quality, BÖ, CL and IW have the lowest fundamental frequency in the acoustic analysis and the two tense voices, CB and IK, seem to have a high fundamental frequency, which correspond to the auditory impression. There are quite big differences in the pronunciation of the r-segment between the speakers in this analysis, depending both on regional and social dialects. The impersonator manages to copy this feature and he often exaggerates the r-segment in order to entertain and strengthen the impression of the voice imitation. In these imitations the texts are created to suit vocabulary and other features of the target speaker. The results in the perception test show that the texts are important for the listener in recognising and judging a voice imitation.

According to the perception test it seems like voice quality, pitch register, intonation (the prosody) and speech style are important features in voice imitations. The impersonator may fail with less important features if he succeeds with more critical features of the target speaker and the listener will get the impression of a successful voice imitation. An imitation can also be like a caricature with a lot of 'overshoots' and with some failures 'undershoots', (Zetterholm 1997) in some features, but it does not seem to affect the overall impression. To strengthen the perceptual impression in voice imitations the impersonator may use characteristic words and phrases of the target speakers. The results in this study indicate however, that even a neutral word 'mobilsvar' appears to be successfully imitated for most of the target speakers. The results, both of the auditory and the acoustic analysis as well as the result of the perception test, indicate the significance of phonetics, both segmental, suprasegmental and extralinguistic features, in voice imitations.

REFERENCES

- Bruce, G. (1998). "Allmän och svensk prosodi" Praktisk lingvistik 16. Department of Linguistics and Phonetics, Lund University.
- Bruce, G. and Gårding, E. (1978) "A prosodic typologi for Swedish dialects" In Gårding, E., Bruce, G. and Bannert, R. (eds) Nordic Prosody: Papers from a symposium, 219-228. Lund: Department of Linguistics.
- Elert, C-C. (1966) "Allmän och svensk fonetik" Norstedts Förlag AB, Stockholm.
- Laver, J. (1994) "Principles of phonetics" Cambridge: Cambridge University Press.
- Pittam, J. (1994) "Voice in Social Interaction. An Interdisciplinary Approach" SAGE Publications, California.
- Stenström, A-B. (1990) "Lexical items peculiar to spoken discourse" The London-Lund Corpus of Spoken English. Description and Research, Jan Svartvik (ed.), 137-176.
- Zetterholm, E. (1997) "Impersonation: a phonetic case study of the imitation of a voice" Working Papers, Department of Linguistics, Lund University 46, 269-287.
- Zetterholm, E. (1999) "Auditory and acoustic analysis of voice quality variations in normal voices" Proceedings of the XIVth International Congress of Phonetic Sciences ICPHS-99, San Francisco, 973-976.
- Zetterholm, E. (2000) fort. "Voice imitation – different ways of saying 'mobilsvar'" Working Papers, Department of Linguistics, Lund University 48.