Intonation Pattern and Duration Differences in Imitated Speech

Elisabeth Zetterholm

Department of Linguistics and Phonetics
Lund University, Sweden
elisabeth.zetterholm@ling.lu.se

Abstract

The present paper is a study of one impersonator and one of his voice imitations. The aim of the study was to investigate how he changes his own voice and speech behavior in order to get close to the target speaker's. A comparison is made between three recordings of the same text material: the target speaker, the voice imitation and the impersonator's own voice and dialect. The results presented in this paper focus on mean fundamental frequency, intonation pattern and duration differences. It is obvious that the impersonator changes his dialect and articulation in the voice imitation.

1. Introduction

Due to anatomical differences there are certain limits when imitating another speaker's voice and speech behavior. The probability of succeeding in producing a close copy of another organically different speaker's utterance is low [5]. According to Laver [5] mimicry is a stereotyping process and that does not involve exactly copying the target speaker. One question is how and to what extent an impersonator changes his own voice and speech behaviour acoustically in order to give the auditory impression of someone else's voice?

To get close to the target speaker and to succeed with the voice imitation the impersonator has to change his own voice and speech behavior in a number of ways. He has to identify important and characteristic features of the target speaker's voice and speech style and know how to change his own voice in order to succeed with the impersonation. Some of the target speaker's characteristic features are related to his or her regional and social dialect, some features are individual phonetic habits. In the voice imitation some important features may be exaggerated and some less important features may be neglected. The audience will still get the impression of a successful voice imitation [6].

In a case study of mimicry, Eriksson & Wretling [4] found that the duration was almost perfectly parallel between the voice imitation and the natural rendition at word level as well as at segmental level, in their recordings. The word durations in the imitations were more similar to the impersonator's own speaking style than to the target speaker. For the study in this paper another impersonator has been recorded.

2. Method

This is a part of a larger study about voice imitation and the results presented in this paper focus on the mean fundamental frequency, intonation pattern and duration differences. A comparison is made between three recordings of the same text material, one recording with the target speaker, one with the voice imitation and one recording with the natural voice of the impersonator.

Three different aspects of the speech have been investigated. First, mean fundamental frequency is compare between the three recordings. Fundamental frequency was measured every 10 ms and the F0 mean calculated. Secondly, the difference in timing between the tonal peak and the VC-boundary is measured in the Swedish accent 1-word *komponerar* (compose). Finally, the speech files were labeled at word level and the words' durations measured. A comparison is made between the three recordings. For comparison, a male speaker recorded the same text material and the word's duration measured.

2.1. Material

Three recordings of the same speech material have been analyzed: One recording with the original speaker taken from public appearances, and two recordings with the Swedish professional impersonator, Anders Mårtensson, one of which was recorded with the voice imitation of the target speaker and one with the impersonator's own natural voice. The recordings by Anders Mårtensson were made particularly for this study, in a studio. Mårtensson has been a professional impersonator for about 10 years. The target speaker was a well-known Swedish TV-personality in the seventies.

The duration of the target text material used for this analysis is approximately 12 s.

2.2. The target speaker

The target speaker was a well-known Swedish TV-personality in the seventies. He lived in Lund in south Sweden and he had a dialect from this area, 1A according to the prosodic typology for Swedish dialects by Bruce & Gårding [3]. The target speaker has a social dialect, which resembles Lunds' academic dialect. The dialect in Lund differs from dialects in the neighborhood, probably due to the influence from the university with students and teachers speaking a number of different dialects. The diphthongs in this dialect are not as marked as they are in other dialects in south Sweden. However, the pronunciation of /r/ is uvular [ʁ], as in other Scanian dialects.

2.3. The impersonator

The Swedish professional impersonator Anders Mårtensson lives in the western part of Sweden and he has a dialect from the transition area between east and west of Sweden, *götamål*, influenced by a central standard Swedish dialect, as described in the prosodic typology for Swedish dialects by Bruce & Gårding [3]. The pronunciation of /r/ is usually with a trill [r] or a retroflex in Mårtensson's dialect.

3. Results

3.1. Mean fundamental frequency

The mean fundamental frequency (F0) is higher in the voice imitation than in the target voice, see Table 1. The impersonator exaggerates the mean fundamental frequency in his efforts to get close to the target voice. It is clear that Mårtensson changes his own mean fundamental frequency in the voice imitation and that he is closer to the target voice than his own natural voice. The overshoot in mean fundamental frequency may depend on the exaggerated tense voice quality in the voice imitation.

Table 1. Mean fundamental frequency (Hz) and standard deviation (Hz) for the three recordings.

	F0 (Hz)	Std dev. (Hz)
Target voice	174 Hz	44 Hz
Imitation	204 Hz	38 Hz
Mårtensson	112 Hz	20 Hz

3.2. Intonation pattern

The F0 peak associated with the stressed syllable is earlier in east and west Swedish dialects than in south Swedish dialects, according to Bruce [1, 2]. In words with accent 1 the peak is before the stressed syllable in east and west Swedish dialects. In south Swedish dialects the peak is later in the stressed syllable. In the recordings used for this study, the accent 1-word *komponerar* (compose) is used five times. The tonal peak is in the beginning of the vowel [eː] in the word *komponerar* (compose), as expected, in the recording of the target speaker with the south Swedish dialect. In the recording of Mårtensson's own voice and dialect from the transition area between east and west of Sweden, the tonal peak is before the stressed vowel, as expected, see Figure 1.

A comparison between the three recordings shows that the impersonator changes his tonal pattern in the voice imitation and uses the tonal pattern of the south Swedish dialect. This is the case in all five occurrences of the word *komponerar* (compose) in this text. The difference in timing between the tonal peak and the VC-boundary (the onset of the stressed vowel) is given in Table 2. E.g. + 19 ms (the target voice) means that the VC-boundary is 19 ms before the tonal peak, and – 19 ms (Mårtensson) means that the VC-boundary is 19 ms after the tonal peak.

Table 2. Timing of F0 peaks (ms) in the word *komponerar* (compose), five occurrences. Values are relative to the VC-boundary.

	1	2	3	4	5
Target voice	+ 19 ms	+ 56 ms	+ 58 ms	+ 60 ms	+ 75 ms
Imitation	+ 12 ms	+ 83 ms	+ 76 ms	+ 66 ms	+ 83 ms
Mårtensson	- 19 ms	- 34 ms	- 34 ms	- 17 ms	- 20 ms

The relation between the VC-boundary and the tonal peak in the word *komponerar* (compose), occurrence number 4, is shown in Figure 1. The intonation pattern is the same for all five occurrences.

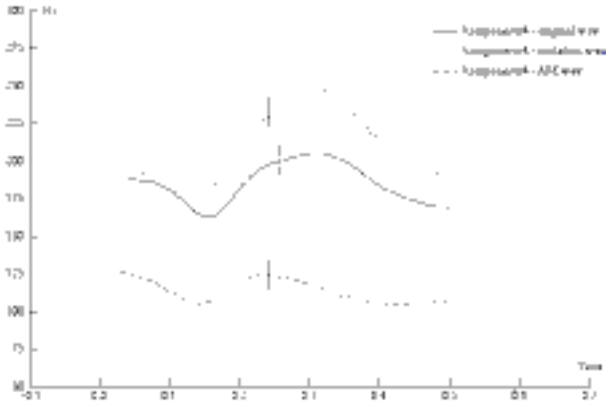

Figure 1. F0 contour of the word *komponerar* (compose) in the three recordings, the target speaker (solid line), the voice imitation (dotted line) and the impersonator's own natural voice (dashed line). The VC-boundary's position is marked with a vertical line.

1.3. Duration differences

1.3.1. The three recordings

There are some differences between the total duration of the three recordings, the target voice, the voice imitation and Mårtensson's own voice. The recording with the voice imitation has the largest duration, 13,5 s. Table 3 shows the total duration and the duration difference between the imitation and the target voice. The duration difference between the voice imitation and the impersonator's own voice is shown in Table 4. The differences between his own voice and the voice imitation may depend on the changes in his vocal apparatus and articulation in order to get close to the target voice.

Table 3. Total duration (s) for the recordings with the target voice and the voice imitation.

	Total duration (s)	Diff. (s)
Target voice	11,89 s	
Imitation	13,58 s	+ 1,69 s

Table 4. Total duration (s) for the recordings with the voice imitation and the impersonator's own voice.

	Total duration (s)	Diff. (s)
Imitation	13,58 s	
Mårtensson	12,42 s	- 1,16 s

The word durations in the three recordings differ from each other. There are differences concerning e.g. the duration of the pauses, which are slightly shorter in the recording with Mårtensson's own voice than in the impersonation. In five cases out of six the word *jag* (I) has a longer duration in the recording of the voice imitation as compared to the two other recordings, see table 5.

Table 5. Duration (s) of the six occurrences of the word *jag* (I).

	1	2	3	4	5	6
Target voice	0,47 s	0,11 s	0,11 s	0,16 s	0,16 s	0,15 s
Imitation	0,63 s	0,12 s	0,26 s	0,42 s	0,13 s	0,23 s
Mårtensson	0,26 s	0,27 s	0,15 s	0,12 s	0,11 s	0,11 s

There is no clear pattern in the duration differences between the three recordings. The durations at word level in the recording with the voice imitation is not closer to the recording with Mårtensson's own voice, in the whole utterance, than the recording with the target speaker. The impersonator changes his duration pattern in his voice imitation. The audible impression is that the target speaker has a specific speech style with a high intensity and speech rate, which is an important characteristic feature of this speaker.

The deviation (in percent) between the target voice, as a reference, and the voice imitation is shown in Figure 2. The difference in the word *jag* (I) is clear, especially in the 4th case. One word is missing in the voice imitation and that explains the 50% faster speech rate in the voice imitation's beginning.

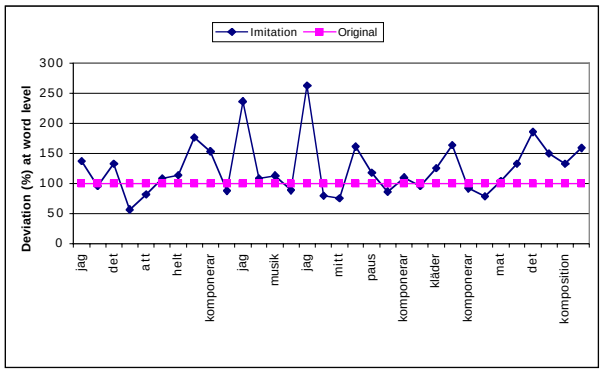

Figure 2. The voice imitation's duration deviation (in %) from the target voice.

The voice imitation's duration deviation (in percent) from the natural voice of the impersonator is shown in Figure 3. There are clear differences between the recordings, e.g. in the word *jag* (I), especially in the 1st and 4th cases. There is a great deviation at the end and the explanation for that may be that the impersonator changes his dialect and uses a diphtongized vowel in his voice imitation.

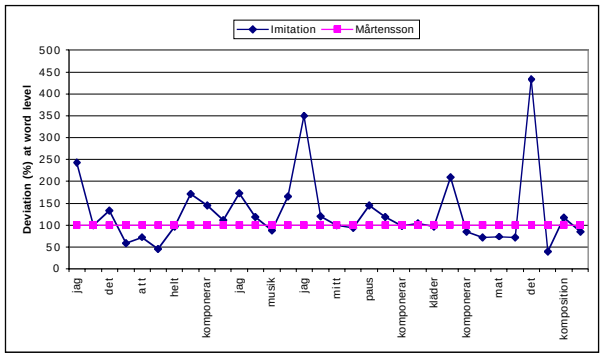

Figure 3. The voice imitation's duration deviation (in %) from the natural voice of the impersonator.

In Figure 4 the voice imitation's and Mårtensson's duration deviation (in percent) from the target voice is shown. In this Figure it is obvious that there is a difference between the voice imitation and the impersonator's own natural voice. The duration pattern is not parallel in the whole utterance.

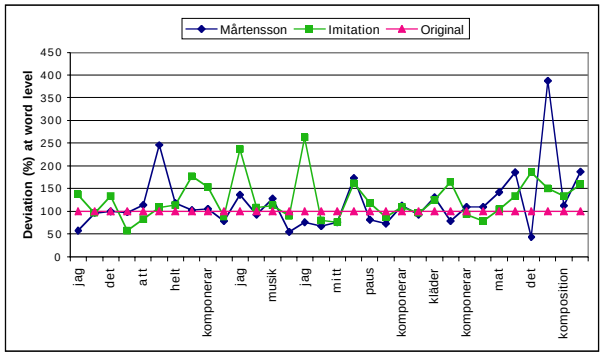

Figure 4. The voice imitation's and Mårtensson's duration deviation (in %) from the target voice.

1.3.2. Recordings of one male speaker

For comparison, and in order to understand how large the duration differences generally are when one speaker read the same text several times, a male speaker was recorded. He read the same text material as the impersonator. The speaker was recorded four times in two days at two different times of the day. The instructions given to him were to simply read the text. He did not know what the recording was going to be used for.

It is obvious that there are only small differences between the four recordings. The total duration of each recording and differences in duration, with the 1st reading as the reference, is shown in Table 6.

Table 6. Total duration (s) for the four recordings with the male speaker.

	Total duration (s)	Diff. (s)
Reading 1	9,08 s	
Reading 2	8,75 s	- 0,33 s
Reading 3	9,33 s	+ 0,25 s
Reading 4	9,92 s	+ 0,84 s

The word duration differences (in percent) between the four recordings are shown in Figure 5. Except for a few exceptions in the 4th reading the timing is very similar and almost parallel in all recordings. The differences in the last reading may depend on the situation and a boring effect. The male speaker was asked if he had tried to read as similar as possible, but that had not been his intention.

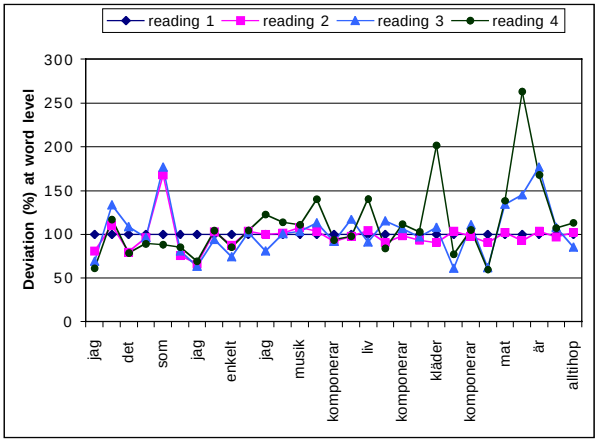

Figure 5. The 2nd, 3rd and 4th reading's duration deviation (in %) from the 1st reading.

2. Discussion

The results indicate that it is possible to get close to another speaker's voice and speech behavior acoustically, despite organically differences between the speakers. It is obvious that the impersonator changes his duration pattern in the voice imitation. The recording of the voice imitation and the recording with the impersonator's own voice do not show the same duration pattern as the target voice. The results presented in this paper indicate that it is possible, for this impersonator, to change his duration pattern at word level in this voice imitation. The voice imitation's duration pattern is not parallel to Mårtensson's duration pattern in the whole utterance. He also changes his mean fundamental frequency, his dialect and intonation pattern, in order to get close to the voice and speech behavior of the target speaker. These changes in his vocal apparatus have effect on the articulatory timing and may explain the duration differences in this voice imitation.

The results in this study do not confirm the results in the study by Eriksson & Wretling [4] concerning speakers' ability to change the duration pattern in voice imitation. In four recordings made by a speaker, who used his own voice, it was shown that the duration pattern changes very little in all readings. Even if we do not say the same thing in exactly the same way every time, the speaker's phonetic habits have a clear effect on the duration pattern. The question is: How individual is the speakers' duration pattern?

3. References

[1] Bruce, G., 1993. Accentuation and timing in Swedish. In *Folia Linguistica Acta Societatis Linguisticae Europaeae, Tomus XVII*. W U Dressler (ed.). Mouton Publishers, 221-238.

[2] Bruce, G., 1998. *Allmän och svensk prosodi*. Praktisk lingvistik 16. Department of Linguistics and Phonetics: Lund University.

[3] Bruce, G.; Gårding, E., 1978. A prosodic typology for Swedish dialects. In *Nordic Prosody: Papers from a symposium*, E. Gårding, G. Bruce, R. Bannert (eds.). Department of Linguistics: Lund University, 219-228.

[4] Eriksson, A.; Wretling, P., 1997. How flexible is the human voice? – A case study of mimicry. In *Proc. Eurospeech '97, Vol. 2*. Rhodes: 1043-1046

[5] Laver, J., 1994. *Principles of phonetics*. Cambridge: Cambridge University Press.

[6] Zetterholm, E., 2001. Voice imitation – different ways of saying *mobilsvar*. In *Working Papers 48*, Department of Linguistics: Lund University, 193-207.