

Dept. for Speech, Music and Hearing
**Quarterly Progress and
Status Report**

**A comparative survey of
phonetic features of two
impersonators**

Zetterholm, E.

journal: Proceedings of Fonetik, TMH-QPSR
volume: 44
number: 1
year: 2002
pages: 129-132



**KTH Computer Science
and Communication**

<http://www.speech.kth.se/qpsr>

A comparative survey of phonetic features of two impersonators

Elisabeth Zetterholm

Dept. of Linguistics and Phonetics, Lund University

Abstract

Do two impersonators select the same features of one target speaker? The present study is a comparative survey of two professional impersonators imitating the same target speakers. It is obvious that the impersonator has to capture many different phonetic features of the target speaker and some individual features seem to be more important for a successful voice imitation and for the listener when recognizing the target voice.

Introduction

To succeed with voice imitation an impersonator has to identify important and characteristic features of the target speaker's voice and speech behaviour. Whatever the aim of the impersonation is, entertainment or mimicing the target speaker as closely as possible, the impersonator has to pick out the most characteristic features. When different impersonators imitate the same target speaker - do they select the same features?

When performing on stage, the goal of the voice imitation is not always to be an exact reproduction of the target speaker. The aim is to entertain and the impersonator often focuses on the most prominent features of the target speaker. Some features may be exaggerated and the voice imitation more like a caricature of the target speaker. On stage, the impersonator uses gestures, body language and other attributes to strengthen the impression of the voice imitation as well.

According to the results of a previous perception test and the comments from the participants (Zetterholm, 2001) it seems that voice quality, pitch register, prosody (intonation and phrasing) and speech style (tempo, rhythm and articulation) are important features to succeed with voice imitations. In that test the listeners were asked to grade and comment on the degree of success of 12 voice imitations made by one impersonator. For the present study recordings made by two professional impersonators imitating six target voices are used.

Method

This is a comparative survey of phonetic features of two impersonators imitating the same

target voices. The voice imitations are made by two Swedish professional impersonators. Some acoustic measurements have been done, such as the mean fundamental frequency for all voice imitations and the impersonators' own voice, the formant frequencies of the i-vowel in one of the voice imitations and the lower edge of the spectrum of /s/ has been measured in two of the voice imitations. An informal listening test has been done by colleagues at the Department of Linguistics and Phonetics at Lund University. In addition to that, two speech therapists at the Department of Logopedics and Phoniatrics at Lund University have made a critical and close analysis concerning the different voices, focusing on the voice quality of the voice imitations as well as the natural voices of the two impersonators. The listeners are familiar with the target voices.

Material

Recordings of voice imitations, made by two impersonators, of six different well-known Swedish voices were available for this study. The texts are different, but related to the target speakers' profession. Four of the target speakers are well-known present or former politicians and two are famous TV-hosts. 10 out of 12 recordings are primarily meant for entertainment purposes. The other two recordings are parts of larger utterances recorded in order to mimic the target voices as close as possible. There are no comparison recordings with the target voices.

The impersonators

The two male Swedish impersonators used in this study have been professional for about 10 years. They do not live in the same area in

Sweden but they both speak a dialect from the transition area between east and west of Sweden, *götamål*, according to the prosodic typology for Swedish dialects by Bruce & Gårding (1978).

The target voices

All six target speakers have a dialect from the southern part of Sweden. Some are from the western and some from the eastern part. There are differences between the dialects concerning both segments, e.g. the r-segment, and the intonation pattern. The target speakers have different dialects and none of them have the same dialect as any of the impersonators. They will be presented with initials only in this paper.

Results

Auditory impression

The colleagues at Lund University has been listening to the voice imitations. There has been a discussion about the general impression and more specific phonetic features of each of the voice imitations. Two speech therapists have completed the analysis of the voice imitations and the impersonators' own voices.

According to the speech therapists, Impersonator I has a sonorous but slightly leaky voice. Impersonator II has a slightly strained and hypernasal voice quality. Both impersonators speak with a normal male pitch level.

The general impression of all voice imitations, according to the phoneticians, is that it is obvious that the impersonators have picked out the most characteristic features of the target speakers. That doesn't mean that the two voice imitations of the same target speaker sound similar. The impression is that both impersonators have the ability to imitate those voices with global success. There are passages, especially in the pronunciation of vowels, which could be improved and make the voice imitation more convincing and closer to the target speaker.

A summary of the comments from the phonetician group and the speech therapists concerning each of the voice imitations will now be presented.

AS: Both impersonators capture the target speakers' speech style with a slow speech tempo, a rather wide pitch range and exaggerated prolonged stressed vowels and a

slightly strained voice quality. The r-segment is an alveolar trill [r], which corresponds to the pronunciation of the target speaker. In spite of these characteristic features, the impression is that the voice imitations are rather different, according to comments from the listeners.

CB: Some of the most characteristic features of this target speaker are his pronunciation of a uvular trilled [ʀ], his hesitation sound in the beginning of a phrase and the way he uses rhetorical pauses. Both impersonators capture these features. In the voice imitations they use a slightly strained and creaky voice quality, especially impersonator I.

HV: The imitations of this target voice are different. One of the group members, who is not familiar with the target voice, was uncertain whether if the impersonators were imitating the same target voice. In spite of that, there are some important characteristic features that both impersonators focus on; an energetic distinct articulation, a speech rhythm like staccato (especially impersonator II), the intonation pattern, the uvular fricative r-segment [ʁ], the characteristic [ʃ]-like s-sound and the less sonorous, slightly strained and creaky voice quality (at least impersonator I).

IK: None of the voice imitations is very close to the target voice, according to the listeners' comments. The impression is that they are more like caricatures. Both impersonators try to capture the same pronunciation of the s-sound, sometimes like a [ʃ], as the target speaker and the tense and hypernasal voice quality. The pitch is high, particularly in the voice imitation by impersonator II, and the voice quality is not sonorous in any of the voice imitations and not close to the target voice.

IW: Both voice imitations are very close to the target speaker and successful, according to the phonetic listeners. The low pitch, the relaxed speech style and the pronunciation of a damped i-vowel [i:] are important features of the target speaker. The speech therapists do not agree that the voice imitations are close to the target voice. A creaky and strained voice quality occurs throughout the voice imitations and impersonator I has a hypernasal voice quality and impersonator II has a hyponasal voice quality, as well.

LO: Impersonator I has captured the dialect and the clearly downstepped intonation pattern of the target speaker. The voice imitation made

by impersonator II is more like an imitation of the west Swedish dialect. The phoneticians mean that there is something in the individual speaking style of the target speaker that is missing in both voice imitations. The impersonators speak with a slightly hypernasal voice quality and a high pitch, according to comments from the speech therapists.

Duration

All texts used in these voice imitations are different. The duration of the texts vary from 9 to 17 seconds, including the texts with the impersonators’ own voices.

F0

In the analysis of mean fundamental frequency it is clear that the impersonators have the same conception about the variation in F0 between the target voices. Impersonator II has a higher mean F0 when speaking with his own natural voice compared to impersonator I, and that may influence the mean F0 in the voice imitations. In four out of six voice imitations impersonator II has a higher mean F0. See Table 1.

Table 1. F0 mean (Hz) and std.dev. for impersonator I and impersonator II speaking with their own voices and the voice imitations.

	Impersonator I		Impersonator II	
	F0 mean	Std. dev.	F0 mean	Std. dev.
Own voice	112	20	127	53
Imitations:				
AS	133	40	142	35
CB	125	23	130	28
HV	91	14	119	28
IK	198	23	255	37
IW	99	16	97	15
LO	142	39	133	25

The voice imitations of IW have the lowest mean F0 and the voice imitations of IK have the highest mean F0, as expected. The high mean F0 in the voice imitation of IK, made by impersonator II, is an imitation of a political speech. He tries to imitate the specific voice quality and speech style used in this context. This is probably the explanation for the very high F0.

There is a rather big difference in mean F0 between the two voice imitations of HV. Impersonator I use a creaky voice quality and that may be related to the low mean F0.

The Swedish vowel [i]

In the voice imitations of the target voice IW a damped vowel [i:] is obvious in the auditory analysis. In a study by Björsten et al. (1999) it was shown that the damped i-vowel has a higher F1 and a lower F2 compared to Swedish standard language. The formant frequencies of some i-vowels in the voice imitations used in this study have been measured. It is obvious that the first formant, F1, is increased and the second formant, F2, is lowered in all the occurrences of the i-vowel in both voice imitations of the target speaker IW.

The long vowel [i:] often has a voiced fricative offglide [i^h], especially in the recording with impersonator I, a feature that characterizes the Swedish damped i-vowel (Björsten et al., 1999, Elert, 1991). Besides that, there is a clear hesitation sound after the vowel, in words like *vi* (we) and *i* (in) in the voice imitation made by impersonator I.

The s-segment

In the voice imitations of HV and IK the audible impression is that some of the s-segments are pronounced more like a [ʃ], which is a characteristic feature of the target speakers. The lower spectral edge of some of the occurrences of the s-segment have been measured in both voice imitations of the target speakers HV and IK.

According to Lindblad (1980) the lower edge of [s] in spectrum is expected to be higher compared to [ʃ]. There are differences, with a lower frequency for [ʃ], between the measurement of the lower edge in spectrum in words with the audible impression of [ʃ] compared to [s]. This is the case in both voice imitations and confirms the listeners’ impression.

Summary and conclusions

All texts in these voice imitations refer to the target voices’ profession. That may influence the listener when recognizing an imitated voice (Zetterholm & Sullivan, 2002). The listener may also have expectations about important and characteristic features of the target speaker’s voice and speech behaviour. 10 out of 12 recordings used for this study are primarily meant for entertainment. Probably the impersonator exaggerates some features to a greater extent when entertaining, compared to a

recording made primarily for a phonetic study. It is hard to say if that is the case in these recordings.

There is a great variation in the pronunciation of the r-segment in Swedish dialects (Elert, 1991). The target speakers in this study represent both alveolar trill (e.g. AS), uvular trill (CB) and uvular fricative (HV). For some of the target speakers this is an important individual characteristic feature. Both impersonator I and impersonator II have captured the different pronunciation of the r-segment in these voice imitations.

In almost all voice imitations the listeners comment that the impersonators have captured the speech style of the target speaker (e.g. AS, CB, HV, IW). It seems that this is one important feature in voice imitation. To capture the prosody (intonation and phrasing) of the target speaker is also important in order to succeed with the voice imitation, according to a previous perception test (Zetterholm, 2001). The phonetic listeners in this study point out the impersonations of CB, IW and LO (impersonator I) as close to the target speakers concerning the prosody.

The auditory impression of the pitch level is confirmed in the acoustic analysis of mean fundamental frequency. The impression is that the pitch level in the voice imitations corresponds to the target voices. The listeners did not comment on the difference in pitch between the voice imitations of HV, but there is a clear difference in mean F0 in the acoustic analysis. The differences between the voice imitations of IK may depend on the speech style, political speech, in the imitation made by impersonator II. The impression of the characteristic pronunciation of a damped Swedish i-vowel [i:] corresponds to the results of the measurements in the acoustic analysis of the voice imitations of IW. The [ʃ]-like pronuncia-

tion of the s-segments is confirmed in the voice imitations of HV and IK, as well.

It seems that the impersonator has to capture many different features of the target speaker, not only the dialect or the pronunciation of one characteristic speech sound, for example, to succeed with the voice imitation. On stage, a global success may be enough, but in a critical listening there are passages which reveal the impersonator. The result of this study indicates that there may be some individual features in a speaker's voice and speech behaviour that seems to be more important when recognizing a voice. The two professional impersonators in this study have picked out and try to capture the same characteristic features for most of the target speakers. Despite that, the voice imitations are different according to the comments from all listeners, the phoneticians and the speech therapists.

References

- Björsten S, Bruce G, Elert C-C, Engstrand O, Eriksson A, Strangert E & Wretling P (1999). Svensk dialektologi och fonetik – tjänster och gentjänster. *Svenska landsmål och svenskt folkliv*, 7-24.
- Bruce G & Gårding E (1978). A prosodic typology for Swedish dialects. In: Gårding E, Bruce G, Bannert R, eds, *Nordic Prosody: Papers from a symposium*. Sweden: Department of Linguistics, Lund University, 219-228.
- Elert C-C (1991). *Allmän och svensk fonetik*. Stockholm: Norstedts Förlag AB.
- Lindblad P (1980). *Svenskans sje- och tje-ljud i ett allmänfonetiskt perspektiv*. Lund: CWK Gleerup.
- Zetterholm E (2001). Voice imitation – different ways of saying mobilsvär. In: Svantesson J-O, ed, *Working Papers 48*. Sweden: Department of Linguistics Lund University, 193-207.
- Zetterholm E & Sullivan K P H (in prep). The impact of semantic expectation on the acceptance of a voice imitation.