The same but different – three impersonators imitate the same target voices

Elisabeth Zetterholm

Lund University, Sweden & Umeå University, Sweden
E-mail: elisabeth.zetterholm@ling.lu.se, elisabeth.zetterholm@ling.umu.se

ABSTRACT

To succeed with the voice imitation, the impersonator has to select the perceptually significant features of the target speaker's voice and speech behavior. Do different impersonators select the same features of a specific target speaker? To try to answer this question, a comparative study was done of phonetic features in voice imitations made by three male Swedish impersonators imitating the same target voices. Auditory and acoustic analyses were conducted of the imitators' natural voices, nine target voices and 22 voice imitations. The results indicate that the impersonators in this study have selected and try to imitate almost the same characteristic features of the target speakers. Both auditory impressions and acoustic measurements, indicate that it is possible to get close to the voice and speech of the target speaker.

The purpose of the study was to obtain an insight into centrally important features of specifically impersonation, which may also give us a general understanding of speaker identification.

1. INTRODUCTION

Recent research has indicated that professional impersonators are successful in their imitations and are able to get close to the voice and speech behavior of the target speakers [1]. To succeed, the impersonator has to identify and imitate the most characteristic features – vocal, segmental and prosodic – of the target speaker's voice and speech behavior. Some of these features are regional and social markers, some are the speaker's individual phonetic habit and speaking style. Exaggeration, sometimes almost like a caricature, is frequent in the imitations since the goal is often to entertain. According to comments from listeners, this does not affect the overall impression, sometimes it strengthens the impression of the target speaker.

The specific research question investigated in this study is: Do different impersonators, when imitating the same target speaker, select the same features of the voice and speech behavior?

2. MATERIAL

Voice imitations made by three male Swedish impersonators were used. Recordings of nine target voices, 22 imitations and the impersonators' own voices were analyzed.

Two male professional impersonators and one amateur were used in this study. None of them live in the same area and they speak different dialects. Impersonator I has a dialect from the west area of Sweden, impersonator II has a dialect from the east area and impersonator III has a neutral Swedish dialect influenced by the intonation pattern of South Swedish. The dialect categories follow Markham (1997) [2].

The target speakers are well-known male Swedish voices, TV-hosts or politicians. They will be presented with their initials only. Four of the target speakers are imitated by all three impersonators. Only impersonator I imitates all voices.

All texts are different and related to the target speakers' profession. The recordings of the target speakers are taken from public appearances and the recordings of the impersonators' own voices and the imitations are made in studios. The duration of the recordings vary between 9 and 33 seconds.

3. METHOD

Several phoneticians at the Department of linguistics and phonetics at Lund University made an informal listening of all recordings. They were familiar with most of the target voices. They were asked to comment on the voice imitations and try to explain which characteristic features of the target speakers the impersonators have selected. There was a discussion about the general impression of the imitations as well as specific features. The listeners focused on phonetic features such as pitch, voice quality, dialectal markers, speech tempo and individual phonetic habits.

In the acoustic analysis the mean F0 values for the target voices, the voice imitations and the impersonators' own voices were measured, and the formant frequencies of the vowel /i/ in some of the recordings.

One of the target voices, AS, imitated by all three impersonators, was selected and a narrow auditory and acoustic analysis as well as a comparison between all recordings were done. Mean F0, F0 range, intonation pattern and articulation rate were measured.

ISBN 1-876346-48-5 © 2003 UAB

4. RESULTS AND COMMENTS

Auditory analysis

The general opinion of all imitations is that the impersonators have selected almost the same and the most characteristic features of the target speakers. Despite that, the voice imitations are different. The general impression is that the impersonators, especially the professionals, have the ability to imitate these voices with success concerning global impression.

In general it seems that, both impersonator I and II are aware of and manage to copy the different pitch levels of the target speakers. Impersonator III has a rather high pitch level and less variation in his imitations. The impression of a pitch level close to the target speaker seems to be important for the acceptance of a voice imitation, according to the listeners.

According to comments from the listeners, impersonator I and II often manage to change their voice quality to get close to the target speakers. In the imitations by impersonator III it is obvious that he does not change his own voice more than to a certain extent so that his own voice quality is audible in most of his imitations. Generally, it seems that it is hard to change voice quality, both laryngeal and supralaryngeal features, and make a copy of another speaker's voice quality all over.

The target speakers represent different Swedish dialects and no one speaks the same dialect as any of the impersonators. There are differences between the dialects concerning both segments and intonation pattern, but only a few will be mentioned here since these are the most obvious differences between the target speakers in this study. A number of different forms of the phoneme /r/ occurs in Swedish. The most common form is the alveolar trill [r] among the majority of the Swedish dialects except for the South dialect, where a uvular trill [ʀ] or a uvular fricative [ʁ] is used. In the dialect of Stockholm, the /ɛː/-vowel is pronounced more like [eː]. A 'damped' i-vowel occurs in some dialects, sometimes as a social variety. A lowered F2 compared to standard Swedish is one acoustic correlate for [ɨ] [3].

Of the target speakers, CB uses a uvular trilled [ʀ] and HV uses a uvular fricative [ʁ]. This segment is a characteristic feature of CB and HV and exaggerated in all imitations. All other target speakers use an alveolar trill [r]. In the voice imitations of CG and GP the Stockholm dialect and the characteristic pronunciation of the phoneme [eː] have been captured in a clear way and close to the target speakers. IW uses a 'damped' i-vowel and both impersonator I and II manage to copy this.

The impression is that all impersonators are aware of and try to copy the speech style – speech tempo, rhythm and pausing – of the target speakers. A very slow tempo, loud extensive breath and many pauses and hesitation sounds are characteristics of GP, a high speech tempo characterize the target speakers LO and MH. Both CB and HV have a committed speaking style with a rhythm like staccato, a lot of pauses between rather short phrases spoken with a fast speech tempo. Some of the imitations are exaggerated to some extent concerning the speech style, e.g. the speech tempo of MH in the imitation by impersonator III.

Acoustic analysis

There are only small differences in mean F0 between the target voices, except from IK (with a high mean F0) an IW (with a low mean F0). Impersonator III has the highest mean F0, 149 Hz, compared to impersonator I with the lowest mean F0, 113 Hz, and impersonator III, 127 Hz, when speaking with their own natural voices. See Table 1.

	Target voices		Imp. I		Imp. II		Imp. III	
	Mean F0	Std. dev.	Mean F0	Std. dev.	Mean F0	Std. dev.	Mean F0	Std. dev.
			113	38	127	53	149	32
AS	128	41	133	40	142	35	145	45
CB	135	35	125	23	130	28	157	56
CG	121	21	122	17	103	11	136	26
GP	126	42	96	36	-	-	139	48
HV	135	36	91	14	119	28	145	76
IK	207	33	198	23	255	37	-	-
IW	107	27	99	16	97	15	-	-
LO	149	28	142	39	133	25	-	-
MH	147	31	202	40	-	-	218	25

Table 1: Mean F0 and std.dev. in all recordings. The target speaker's initials are in the leftmost column.

The mean F0 of the impersonators' natural voices are reflected in the imitations. In all cases impersonator III has the highest mean F0 and in six out of nine voice imitations impersonator I has the lowest mean F0.

It is obvious that the impersonators change their own mean F0 in order to get close to the target speakers and some of them are rather close. It is clear that especially impersonator I and II have the same conception about the variation in F0 between the target voices. There is less variation between the different imitations made by impersonator III, except for his imitation of MH. The acoustic values correspond to the auditory impression.

There is a difference in the imitations of GP, where impersonator I has a much lower mean F0 while impersonator III has a higher mean F0 compared to the target speaker. The target speaker IW has a low mean F0 and both impersonator I and II imitate this voice with a low mean F0, rather close to the target voice. Corresponding to the auditory impression, the acoustic analysis shows that the mean F0 of the imitations of MH is exaggerated by both impersonator I and III. In the imitations of HV there are obvious differences between the three imitations.

The auditory impression of a 'damped' i-vowel both for the target speaker and the imitations of IW is confirmed in the acoustic analysis. The different occurrences of /i/ were measured and even though the texts are different, there is a clear tendency to a lowered F2.

4. ONE TARGET SPEAKER

A closer comparison was done between the target voice and the imitations of AS. This voice was selected since AS was imitated by all three impersonators and the recordings comparable concerning the recording quality and the content of the text.

Both impersonator I and II try to imitate the audible impression of a sonorous and slightly nasal voice quality of the target speaker. In the imitation by impersonator III his own voice is audible throughout the imitation.

All impersonators have captured, and to some extent exaggerated, the clear articulation with a trilled [r], which is also confirmed in the spectrograms. Concerning the pronunciation of the i-vowels it is obvious that there is no audible 'damped' i-vowel in any of the imitations, which correspond to the impression of the recording with the target speaker. However, the acoustic analysis of the formant frequencies show a lowered F2 in some words in the imitation by impersonator I.

The target speaker has a speaking style with emphasis and a characteristic intonation pattern with stressed syllables and a slow speech tempo. This intonation pattern is copied by impersonator I and II, but not by impersonator III. The F0 range is measured in all recordings and seem to be similar, about 80-245 Hz, which is a rather wide F0 range in normal speech. The same intonation pattern with recurring large F0 excursions for successive accented words without much declination is obvious in the spectrograms of AS himself and impersonator I and II. In the spectrogram of the imitation by impersonator III there is instead a successive narrowing in F0 range of the corresponding intonation pattern. Probably he tries to imitate the emphatic speech style, but performed in a different way. See Figures 1-4.

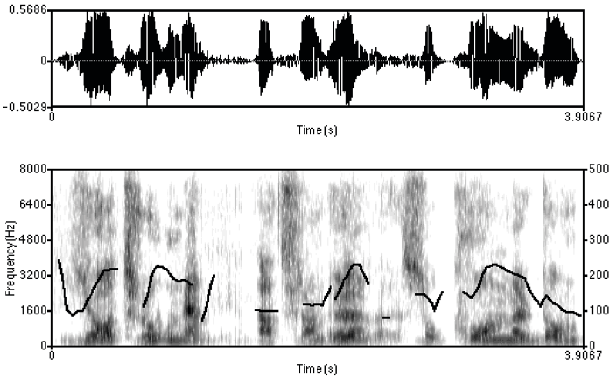

Figure 1: Waveform, spectrogram and F0 contour of the target speaker AS.

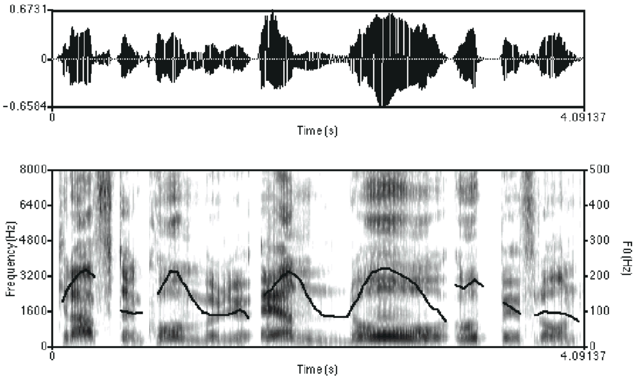

Figure 2: Waveform, spectrogram and F0 contour of the imitation of AS, by impersonator I.

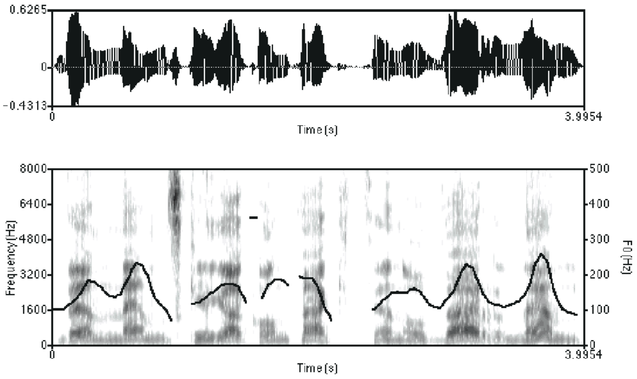

Figure 3: Waveform, spectrogram and F0 contour of the imitation of AS, by impersonator II.

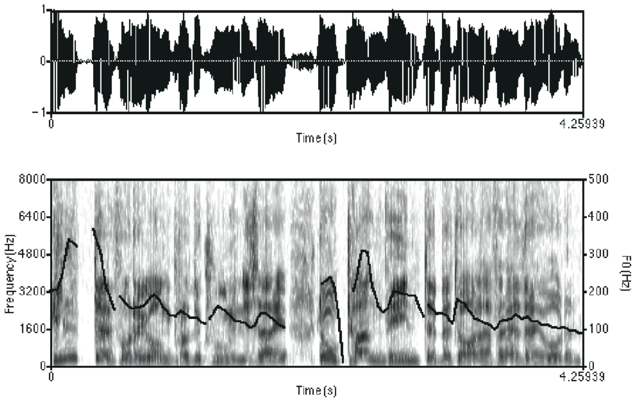

Figure 4: Waveform, spectrogram and F0 contour of the imitation of AS, by impersonator III.

The mean articulation rate for the target speaker is 4,3 syllables per second. The auditory analysis gives an impression of a slower speech tempo. The slow speech tempo is exaggerated in two of the imitations and the mean articulation rate is 2,8 syllables per second in the imitation by impersonator I and 3,0 syllables per second in the imitation by impersonator II. Impersonator III has a

ISBN 1-876346-48-5 © 2003 UAB

slow speech tempo in the first phrase of his imitation, 3,4 syllables per second, but speeds up the tempo in the other phrases and the mean articulation rate is 4,7 syllables per second. The resulting imitations correspond the auditory impression.

4. DISCUSSION AND CONCLUSIONS

It is obvious that the voice and speech imitations by these three impersonators are different, but it is still possible to make a clear identification of the imitated person. When comparing the imitations with the recordings of the target speakers, the listeners agree that the impersonators have selected prominent features of each speaker. According to the listeners, they also try to focus on the same characteristics of each target voice.

The general auditory impression, according to the listeners, is that the impersonators have captured the pitch level, the dialect and the speech style – speech tempo, rhythm, articulation and intonation pattern – as well as individual characteristic features such as hesitation sounds and loud breathing of the target speakers. Some of these features are exaggerated to some extent. In two of the imitations of AS the slow speech tempo is exaggerated. All impersonators try to imitate the emphatic speech style of this speaker and the two professional impersonators manage to copy the intonation pattern of AS, while the amateur perform this speaking style in another way, with declinations.

The acoustic measurement of mean F0 corresponds to the auditory impression, and indicates that the impersonators change their own mean F0 to get close to the target speakers. The results show that the two professional impersonators, called I and II, change their own mean F0 to a greater extent than the amateur, impersonator III. It is also noticeable that the impersonators' natural mean F0 seems to be reflected in the voice imitations. According to the listeners, F0 plays an important role for the acceptance of a voice and speech imitation.

The variation of different voice qualities in normal voices is hard to describe, but the listeners are able to tell if the voice quality in the imitation is close to the target speaker or not. There are passages in the imitations where the natural voice of the impersonator is audible. The two professional impersonators seem to be more successful in changing their own voice quality in order to try to get close to the voice quality of the target speaker, compared to the amateur impersonator. This result may indicate that voice quality is one feature in the human voice that is hard to change.

All three impersonators are able to imitate the different dialects with regional and social markers, e.g. the different pronunciation of the phoneme /r/, the vowel [eː] and the 'damped' i-vowel. The auditory impression of a 'damped' i-vowel is confirmed in the acoustic analysis of the target voice and the imitations of IW.

All texts in these voice imitations deal with topics related to the target speakers' profession. The semantic information may influence the listener when recognizing an imitated voice. This suggestion corresponds to the result in recent research about listeners' expectation and acceptance of an imitated voice [4, 5].

The results of this study indicate that there may be some individual features in a speaker's voice and speech behavior that seem to be more important than others for the recognition of a voice, both considering the features selected by the three impersonators and the comments from the listeners. That may give a clue about individual features useful in a speaker identification task. Moreover, they may also give an insight into what features in the human voice that are hard to change.

ACKNOWLEDGEMENTS

This writing of this paper 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.

REFERENCES

[1] E. Zetterholm, *Voice imitation. A phonetic study of perceptual illusions and acoustic success*, Lund, Travaux de l'institut de linguistique de Lund 44, Lund Universtity, 2003.

[2] D. Markham, *Phonetic Imitation, Accent, and the Learner*, Lund, Travaux de l'institut de linguistique de Lund 33, Lund University Press, 1997.

[3] S. Björsten, G. Bruce, C-G Elert, O. Engstrand, A. Eriksson, E. Strangert and P. Wretling, "Svensk dialektologi och fonetik – tjänster och gentjänster," *Svenska landsmål och svenskt folkliv, Swedish dialects and Folk Traditions*, pp. 7-24, 1999.

[4] K.P.H. Sullivan, E. Zetterholm, J. van Doorn, J. Green, F. Kügler and E. Eriksson, "The effect of removing semantic information upon the impact of voice imitation", in *P roceedings of SST2002*, Melbourne, Australia, December 2002, pp. 291-296.

[5] E. Zetterholm, K.P.H. Sullivan, J. Green, E. Eriksson and P.E. Czigler, "Imitation, expectation and acceptance: the role of age and first language in a Nordic setting", in *Proceedings of ICPhS 2003*, Barcelona, Spain, August 2003.