Acoustic Analysis of Imitated Voice Produced by a Professional Impersonator

Tatsuya Kitamura (t-kitamu@konan-u.ac.jp)
Konan University, Kobe, Japan

Questions & Objective

- Which acoustic characteristics do impersonators adjust to those of the target voice?
- How do they control their speech production system while imitating a voice?

→The answers may offer fruitful insights into studies on the perception of speaker individuality and voice conversion in synthetic speech.

Objective
To explore the acoustical characteristics that a professional impersonator changes from his natural voice to imitate a target voice.

1

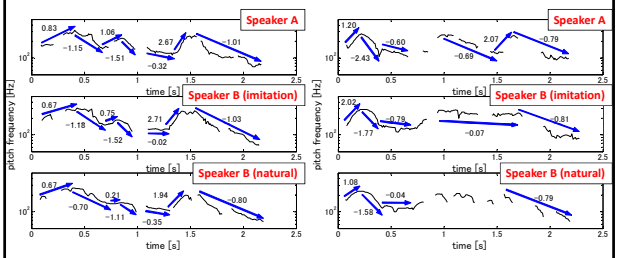
Speech data

- Speakers
  - a professional comic story teller (target speaker, aged 70)
  - a professional impersonator (in imitated voice and his natural voice)
- Sentences (two Japanese sentences)
  - **Sentence 1**: Ichidode i:kara mitemitai, nyo:boga hesokuri kakusutoko. (Just once, I wish to catch my wife hiding her secret savings.)
  - **Sentence 2**: Dekakeru nekoni yukisaki kikeba, ryoko:ga sukide mata tabida. (Asked where to go, the cat replied that he is going on a travel again because he loves to.)

← This Japanese sentence has a humorous play on words.

2

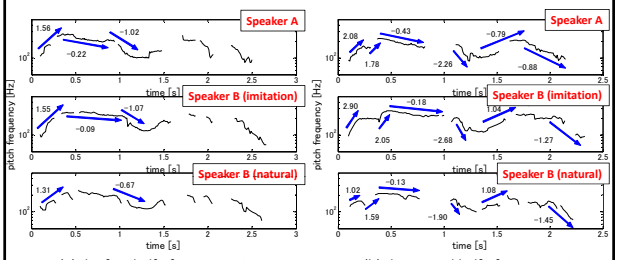
Pitch frequency contour



(a) the first half of Sentence 1 (b) the second half of Sentence 1

Figure 1: Pitch frequency contours of Sentence 1.

3



(a) the first half of Sentence 2 (b) the second half of Sentence 2

Figure 2: Pitch frequency contours of Sentence 2.

The shape and tilt of the pitch frequency contours of the imitated utterances are similar to those of the target ones.

4

Mean pitch frequency

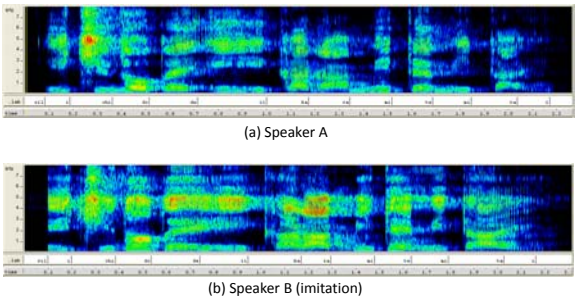
Table 1. The mean pitch frequency of Sentences 1 and 2 in Hz.

	Sentence 1	Sentence 2
Speaker A	167.2	162.2
Speaker B (imitation)	185.1	187.3
Speaker B (natural)	152.0	154.8

approx. 20 Hz higher

5

Spectrogram

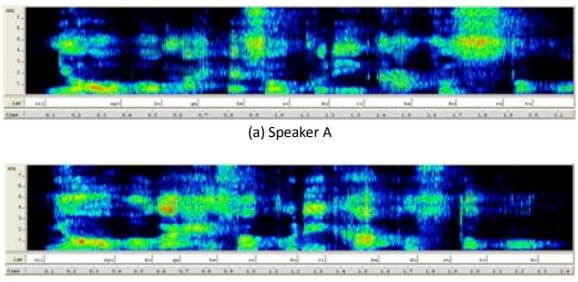


(a) Speaker A

(b) Speaker B (imitation)

Figure 3: Spectrogram of the first half of Sentence 1.

6



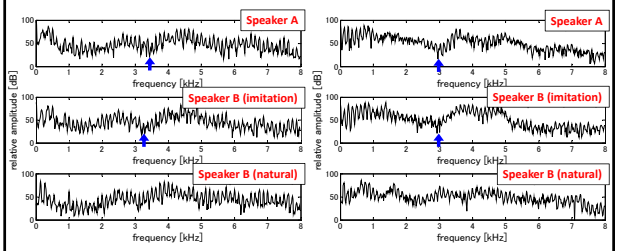
(a) Speaker A

(b) Speaker B (imitation)

Figure 4: Spectrogram of the second half of Sentence 1.

7

Spectrum



(a) long vowel /i:/ in Sentence 1 (b) vowel /a/ in Sentence 1

Figure 5: Speech spectra of vowels in Sentence 2.

Speaker B adjusted not only spectral peaks but also dips.

8

Formant frequencies

Table 2. The first, second, third, and fourth formant frequencies in Hz of the long vowel /i:/ in Sentence 1.

	F1	F2	F3	F4
Speaker A	390.1	2687.5	3171.9	4031.2
Speaker B (imitation)	406.2	2468.8	---	4049.9
	+62 %			
Speaker B (natural)	250.0	2484.4	2906.2	4031.2

differences is less than 4 % except for F2

Table 3. The first, second, third, and fourth formant frequencies in Hz of the vowel /a/ in Sentence 1.

	F1	F2	F3	F4
Speaker A	765.6	1484.4	3578.1	4000.0
Speaker B (imitation)	796.9	1343.8	3593.8	3968.8
	+19 %	-12 %	+17 %	
Speaker B (natural)	671.2	1531.2	3078.1	4140.6

Speaker B must change the vocal tract shape during imitation.

9

Difference between the amplitude of the first and second harmonics (H1-H2)

Table 4. H1-H2 in dB of the long vowel /i:/ and the vowel /a/ in Sentence 1.

	/i:/	/a/
Speaker A	-8.03	-2.22
Speaker B (imitation)	-8.12	-11.55
Speaker B (natural)	3.82	18.3

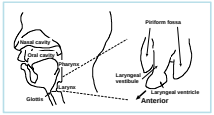
the signs of the values are identical (negative)

Speaker B adjusts the glottal source characteristics to those of Speaker A.

10

Discussion

- The mean pitch frequencies of the imitated utterances are approximately 20 Hz higher than those of the target voice.
  - Speaker B probably exaggerates Speaker A's high-pitched voice.
  - Zetterholm (2001) reported similar results.
- Speaker B imitates the shape of the pitch frequency contour in many parts of his utterances.
  - Akagi and Ienaga (1997) showed that the dynamics of the pitch frequency contour as well as the mean contributes the perceptual speaker identification.
- Speaker B also imitates the shape of the speech spectra.
  - Speaker B controlled the vocal tract shape.
  - Potential source of the dip in the frequency region from 3.0 to 3.4 kHz is the piriform fossae.
- Speaker B changes the glottal source characteristics (H1-H2) during imitation.
  - Speaker B tried to imitate Speaker A's hoarse voice.



11

Conclusion

- The impersonator controlled
  - the mean and dynamics of the pitch frequency,
  - the vocal tract acoustic characteristics,
  - and the glottal source characteristics

  when imitating the target voice.

  → These characteristics are probably important to perceptual speaker identification.
- It should be noted that the results depend on the combination of the target speaker and impersonator in this study.

This research was partly supported by SCOPE (071705011) of the ministry of Internal Affairs and Communications, Japan.

12