Usefulness of inverse estimation using a vocal tract mapping interface for estimating articulatory behavior

Kumamoto University

Kohichi Ogata, Kento Yamamoto, and Masami Ito

Faculty of Advanced Science and Technology, Kumamoto University, Japan

1. Background

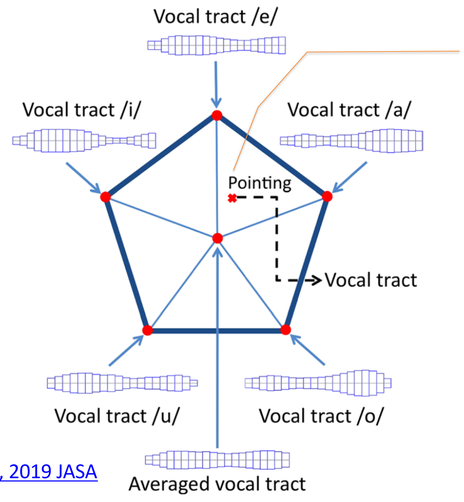
To effectively describe the shape of an entire vocal tract with fewer parameters, a vocal tract mapping interface was developed [1]. Inverse estimation is a long-standing problem in speech production [2-5]. Inverse estimation of the vocal tract shape based on a vocal tract mapping interface was proposed [6].

This study: Whether inverse estimation based on the interface can capture articulatory behavior such as differences in movement timing.

2. Overview of mapping interface

Functions: Synthesis [1], Inverse estimation[6]

- describe the shape of an entire vocal tract with fewer parameters
- generate various vocal tract shapes corresponding to arbitrary points on the interface window
- obtain vocal tract shapes from formant frequencies by inverse estimation

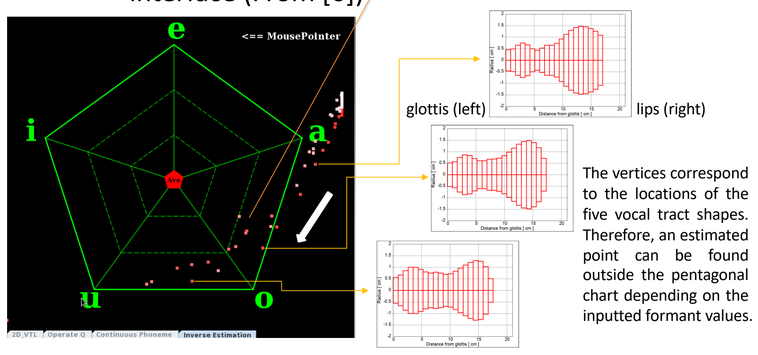


A vocal tract is symbolized as a point

The trajectory of the points depicts the behavior of the vocal tract

See Ogata et al., 2019 JASA for details.

Concept of the vocal tract mapping interface (From [6])



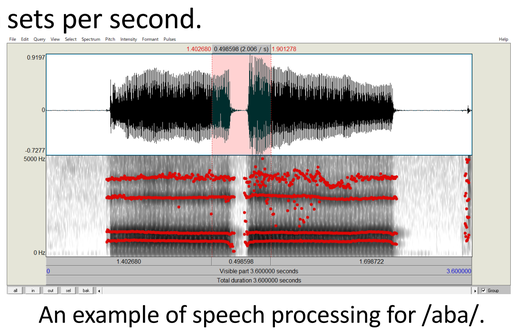
The vertices correspond to the locations of the five vocal tract shapes. Therefore, an estimated point can be found outside the pentagonal chart depending on the inputted formant values.

The trajectory pattern consists of points on the mapping interface window corresponding to estimated vocal tract shapes during the utterance /aba/ [7].

3. Method

3-1. Data acquisition and processing

- A digital recorder (Marantz PMD671) with a microphone (SONY ECM-77B) was used.
- The sampling frequency of the digital recorder was 48 kHz.
- Speech sound files were processed using Praat, and the time sequence of the formant frequencies was obtained as 160 data sets per second.

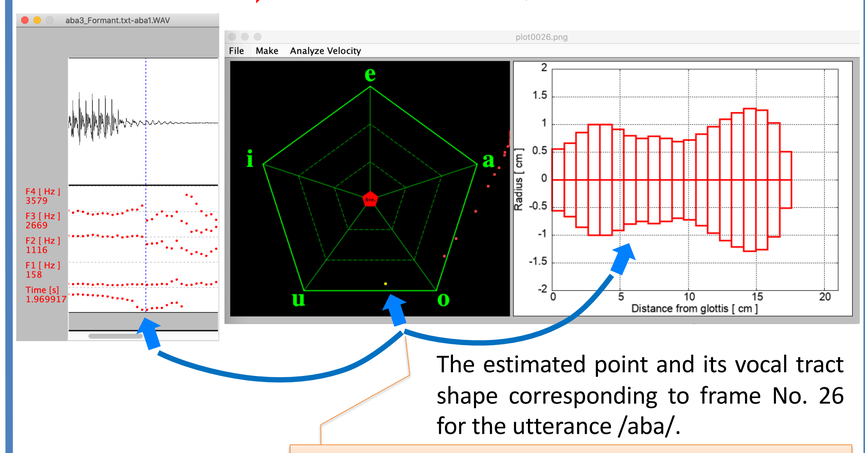


An example of speech processing for /aba/.

3-1. Example of inverse estimation

Inverse estimation — Search the map for target formant frequencies

Formant → Vocal tract shape

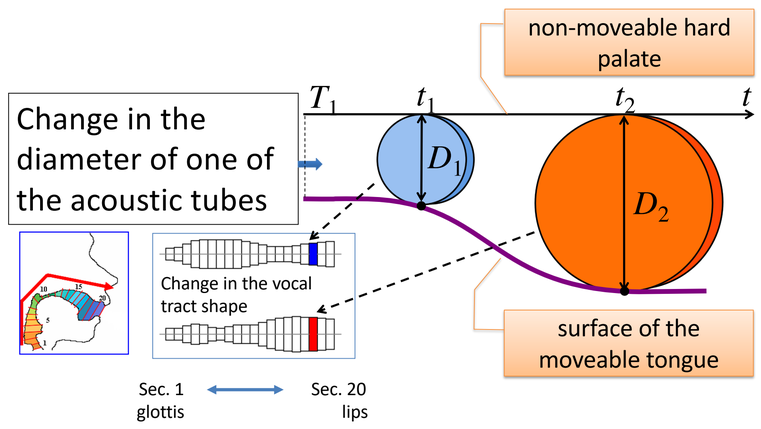


The estimated point and its vocal tract shape corresponding to frame No. 26 for the utterance /aba/.

Corresponding to the same frame number

3-2. Definition of pseudo velocity

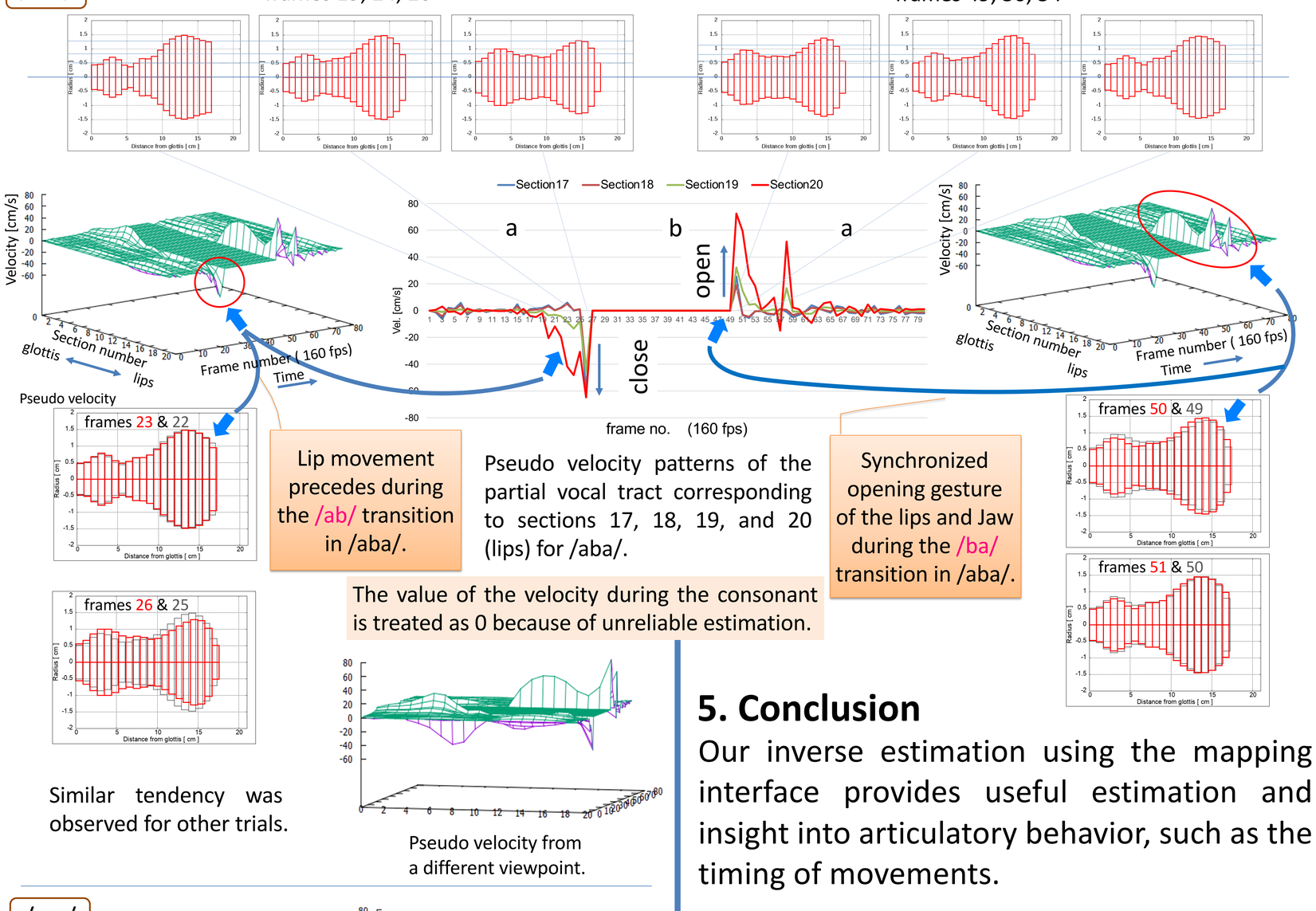
The change in the diameter of each acoustic tube as a function of time was treated as the pseudo velocity of articulatory movement for the sake of simplicity.



4. Results

As an example, the result for /aba/ is shown as follows. Section 20 (red) corresponding to the lip area has an earlier change in velocity than the other sections. This suggests that the lips move before the jaw to close the mouth as quickly as possible for the /ab/ transition in /aba/. In contrast, a synchronized opening gesture can be seen in the four sections at the /ba/ transition.

/aba/



Lip movement precedes during the /ab/ transition in /aba/.

Pseudo velocity patterns of the partial vocal tract corresponding to sections 17, 18, 19, and 20 (lips) for /aba/.

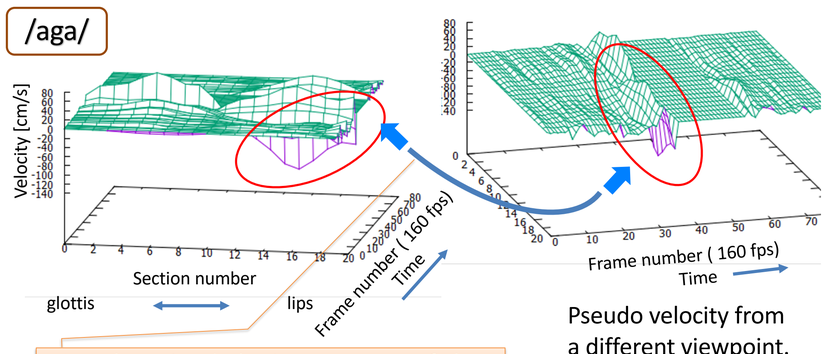
Synchronized opening gesture of the lips and Jaw during the /ba/ transition in /aba/.

The value of the velocity during the consonant is treated as 0 because of unreliable estimation.

Similar tendency was observed for other trials.

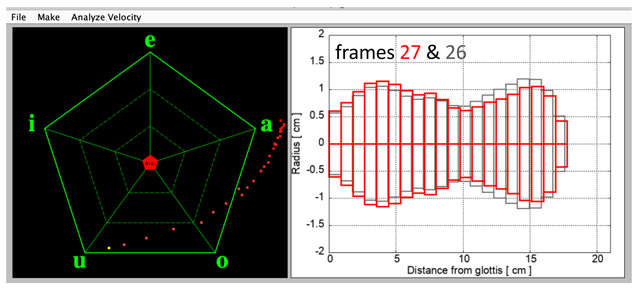
Pseudo velocity from a different viewpoint.

/aga/

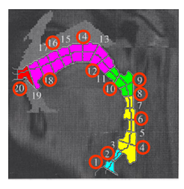


Synchronized closing gesture of the lips and Jaw before /g/ in /aga/.

Pseudo velocity from a different viewpoint.



The estimated points and their vocal tract shapes for the utterance /aga/.



FYI
(another subject)

5. Conclusion

Our inverse estimation using the mapping interface provides useful estimation and insight into articulatory behavior, such as the timing of movements.

Future work

- Analysis of more data, comparison etc.
- Try to estimate the vocal tract shape during consonants

References

[1] K. Ogata and K. Yamashita, "One-click vocal tract mapping interface and its application to signal conversion," IEICE Trans. **J96-A**, 529-540, 2013 (in Japanese).

[2] P. Mermelstein, "Determination of the vocal-tract shape from measured formant frequencies," J. Acoust. Soc. Am., **41**, 1283-1294, 1967.

[3] M. R. Schroeder, "Determination of the geometry of the human vocal tract by acoustic measurements," J. Acoust. Soc. Am., **41**, 1002-1010, 1967.

[4] B. H. Story, "Technique for tuning vocal tract area functions based on acoustic sensitivity functions," J. Acoust. Soc. Am., **119**, 715-718, 2006.

[5] T. Kaburagi, "A method for estimating vocal-tract shape from a target speech spectrum," Acoust. Sci. Technol., **36**, 428-437, 2015.

[6] K. Ogata, T. Kodama, T. Hayakawa, and R. Aoki, "Inverse estimation of the vocal tract shape based on a vocal tract mapping interface," J. Acoust. Soc. Am., **145**, 1961-1974, 2019.

[7] K. Ogata and T. Tanaka, "Inverse estimation of the vocal tract shape from speech sounds including consonants using a vocal tract mapping interface," Proc. 23rd ICA, Germany, 6887-6894, 2019.

Acknowledgment : Part of this work was supported by Grant-in-Aid for Scientific Research JP17K06464 from the Japan Society for the Promotion of Science.