Estimates for the measurement and articulatory error in MRI data from sustained vowel production

Daniel Aalto[1,2], Jarmo Malinen[3], Martti Vainio[2], Jani Saunavaara[4], Pertti Palo[3]

[1]Aalto University School of Electrical Engineering, Department of Signal Processing and Acoustics, [2]University of Helsinki, Institute of Behavioural Sciences,
[3]Aalto University School of Science and Technology, Department of Systems analysis and Mathematics, [4]Medical Imaging Centre of Southwest Finland

Introduction

MRI is widely accessible, safe and efficient way of obtaining full anatomic pictures (3D) of the vocal tract (VT) [2]. For high-quality pictures, the subject have to keep the VT steady in the continuous sampling paradigm. Due to gravity (supine position) and decrease in lung volume, the subject adjusts the production [3,4,6]. Our aim at estimates for the precision of the VT data to validate a computational model for VT resonances using simultaneously recorded MRI data and sound [1].

We study the effect of production time and bite-block on the stability of the VT and formants. We hypothesize:

1) Longer production time results in more movement,

2) Bite-block reduces all movement.

Experimental setting

Figure 1.Siemens Magnetom Avanto 1.5T scanner (left) and sound collecting equipment (right).

vowel		duration		bite-block		recording site
[ɑ] [æ]	x	7 s 18 s	x	+ -	x	MRI Anechoic

We made 2x2x2x2 recordings for one test subject. The two durations correspond to MRI Turbo sequencing durations with 1.8 and 1.2 mm isotropic voxel sizes respectively.

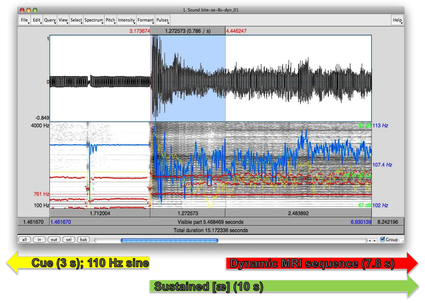

Figure 2. MRI recording session. Sine cue prompts the subject and a dynamic MRI sequence with 5 Hz sampling frequency.

Stability of the vocal tract

The nominal resolution of the MR image was 1.4 mm. This improves because of the gray-scale valued pixels. With bite-block the mandibular opening was constant and for 45 independent measurements (consecutive frames) the standard deviation was 0.4 mm, the true resolution. The tongue tip moves anteriorly and then varies around 65 mm. During the beginning of the sequence, there is more variation: The first 22 frames have a SD of 1.7 mm while the last 22 frames have a SD of 0.8 mm.

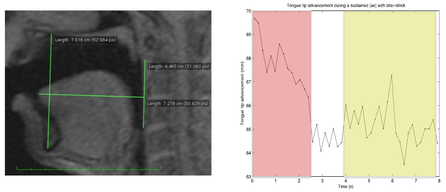

Figure 3. The green lines at left illustrate the measurements “Tongue tip advancement” (TTA) and “Mandibular opening” (MO). At right, the time evolution of TTA during a 8 s [æ] while subject used bite-block.

Stability of the formants

Based on recordings in the anechoic chamber the formants are calculated (LPC). The stability is measured with standard deviation over 22 consecutive frames, compatible with the MRI sequence: std for F_1 was 11 Hz (0.28 st); F_2: 27 Hz (0.33 st); F_3: 75 Hz (0.54 st). The formant trajectories of two productions are depicted in the Figure 4.

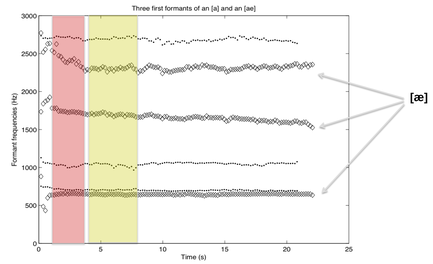

Figure 4. Formant trajectories of [ɑ] and [æ] recorded in an anechoic chamber, no bite-block, very long production time.

The role of the bite-block

A bite-block was put between the incisors to reduce movement artifacts. Based on the acoustic analysis of the anechoic chamber recording, only F_2 sample had smaller SD. The other formants seem to be less stable (the effects are not statistically significant here).

Table 1. Standard deviations of formants.

Means of standard deviations (Hz)		Frames 1-22	Frames 23-44	Last 22 frames
Bite-block -	F_1	21.2	6.55	4.57
	F_2	50.5	26.2	24.3
	F_3	90.8	32.6	28.9
Bite-block +	F_1	10.4	11.9	8.89
	F_2	38.8	12.7	11.2
	F_3	116	70.4	109

The role of the duration

The standard deviation in a given 4 s window of the formants was in average smaller for the longer duration ($F=4.7$, $p<.05$).

Table 2. Standard deviations of formants.

Means of standard deviations (Hz)		Frames 1-22	Frames 23-44	Last 22 frames
Long (μ = 11 s)	F_1	8.2	10.8	7.6
	F_2	52.5	23.2	21.3
	F_3	145	77.8	112
Very long (μ = 21 s)	F_1	23.3	7.7	5.8
	F_2	36.8	15.8	14.1
	F_3	62.0	25.2	25.7

Estimates for the data quality

After the onset adaptation we can expect the following:

•10 s prolonged vowel phonation with onset adaptation time of 3 s before data recording, 7 s sequence

• nominal voxel size 1.8 mm, true resolution 0.5 mm

• stability of the tongue tip position 0.8 mm

• formant stability 0.5 semitones

• 10 Hz ≈ 0,3 mm for stability in F_2/ tongue tip position

Conclusions

First, a fast MRI sequence (7 s) is sufficient for 3D submillimeter data. The stability of sustained vowel production is more limiting than the measurement equipment.

However, our results suggest that longer productions are planned and executed more carefully and hence fluctuation is reduced in terms of standard deviations of formants.

Since formant fluctuation is roughly constant in semitones, the auditory feed-back loop seems to have an important role in the VT motor control. At the beginning of the production, a transition occurs possibly due to gravity; compare with [3,7]. The time scale (a few seconds) is in accordance with [5] assuming the feed-back loop is of order 100 ms.

References

1. Aalto, D., Aaltonen, O., Happonen, R.-P., Malinen, J., Palo, P., Parkkola, R., Saunavaara, J., Vainio, M. Recording Speech Sound and Articulation in MRI. *Proc. Biodevices, Rome, 2011*.
2. Crary, M. A., Kotzur, I. M., Gauger, J., Gorham, M. Burton, S. 1996. Dynamic Magnetic Resonance Imaging in the Study of Vocal Tract Configuration. *Journal of Voice.* 10 No. 4, 378–388.
3. Engwall, O. 2003. A revisit to the application of MRI to the analysis of speech production—testing our assumptions. *Proc of 6th International Seminar on Speech Production.* 43–48.
4. Houde, J., Jordan, M. 1998. Sensorimotor adaptation in speech production. *Science*, 279, 1213.
5. MacDonald. E. N., Purcell D. W., Munhall, K. G. 2011. Probing the independence of formant control using altered auditory feedback. *J. Acoust. Soc. Am.* 129(2), 955–965.
6. Yanagihara, N., Koike, Y. 1967. The Regulation of Sustained Phonation. *Folia phoniat.* 19, 1–18.
7. Stone, M., Stock, G., Bunin, K., Kumar, K., Epstein, M., Kambhamettu, C., Li, M., Parthasarathy, V., Prince, J. 2007. Comparison of speech production in upright and supine position. J. Acoust. Soc. Am. 122 (1), 532–541.

Acknowledgments

The project is supported by Finnish Academy grant no. Lastu 135005.

For further information

Please contact daniel@iki.fi