Vocal Tract Length during Speech Production

Sorin Dusan

Center for Advanced Information Processing, Rutgers University
Piscataway, New Jersey, 08854, U.S.A.
sdusan@caip.rutgers.edu

Abstract

It is known that formant frequencies are inversely proportional with the vocal tract length of the speaker. Although it was observed that vocal tract length of a speaker is variable during speech production, the extent of this variability has not been fully examined in the literature. This paper presents a statistical analysis of the vocal tract length of a female speaker during the production of ten sentences in French. In addition, this paper examines various correlations between vocal tract length, lips protrusion, and larynx height, on one side, and the parameters of Maeda's articulatory model, on the other side. The paper proposes a linear regression model of the vocal tract length as a function of eight articulatory parameters and provides a discussion on the role of the lips and larynx height maneuvers in optimizing the production of speech in terms of achieving high phonetic contrast, high speed, and minimum energy.

Index Terms: vocal tract length, formant frequencies, articulatory correlations, speech optimization

10.21437/Interspeech.2007-17

1. Introduction

It is known that the resonant frequencies of a uniform acoustic tube are inversely proportional with the length of the tube. The acoustic theory of speech production, [1], shows that formant frequencies are lower if the vocal tract length defined between lips and glottis (VTL) of a speaker is greater. Radiographic images of the vocal tract of a Russian male speaker showed that the VTL parameter varied between 16.5 cm and 19.5 cm in six isolated Russian vowels [1]. A detailed examination of the vocal tract of a male speaker producing continuous speech utterances in English suggested that the actual vocal tract length during speech production is determined by two main components, often acting together: lip spreading/protrusion and larynx height [2].

A study based on data from French and Mandarin native speakers found that constrained lip spreading/protrusion induced compensatory vertical larynx displacements and suggested that these compensatory articulations can be explained by a putative contrast-preserving strategy in speech production [3]. This hypothesis, however, was contradicted by a subsequent study which did not find compensatory larynx height movements in three English speakers producing isolated vowels with different amounts of lip protrusion [4]. The hypothesis that larynx depression compensates for labial undershoot was also rejected in [5], which suggested that, in fact, larynx depression is complementary to lip rounding.

The effect of raising/lowering the larynx by 1.5 cm on vowel formant frequencies has been examined in [6], comparing results from two Swedish speakers with results from a model experiment based on the Russian area functions presented in [1]. General agreements have been found, although some differences were attributed to the fact that humans may also change the position of other articulators when they try to modify the position of the larynx. This last point is supported by a more recent study which suggests that the investigation of larynx height position during speech must be embedded in investigations of other articulators involved in shaping the vocal tract [7].

It was suggested that speakers may vary both shape and length of the vocal tract in a coordinated manner in order to achieve desired acoustic targets and that changes in length may reduce additional shape distortion [8], [2]. However, a complete evaluation of the length of the vocal tract in correlation with the other articulators during speech production has not been published. This paper presents an analysis of the variability of the VTL parameter of a female speaker and correlations between the VTL parameter and other articulatory parameters during the production of continuous speech. The paper also emphasizes the special role of the lips and larynx height maneuvers and suggests that these articulators help to optimize the speech production process in order to increase the speed and decrease the energy of phonological constructions.

2. Method

2.1. Data

The data used in this study originates from a study involving a French female speaker (PB) uttering 10 sentences [9]. These data contain digitized midsagittal vocal tract profiles obtained from radiographic films and labial films with a rate of 50 frames per second. The data contain 519 vocal tract profiles (frames) of this speaker. These data were used by Maeda [10] to build an articulatory model based on factor analysis. A version of Maeda's articulatory model [11] was used in this study to represent each vocal tract profile as a function of eight articulatory parameters: jaw (j), tongue body (b), tongue dorsum (d), tongue tip X position (tx), tongue tip Y position (ty), lip X position (lx), lip Y position (ly), and larynx height (lh). The actual articulatory data used in the current study are represented by the 519 parameterized vocal tract profiles by Maeda's model as described in [11]. The eight articulatory parameters of Maeda's model are normalized to have zero mean and unitary variance and standard deviation. The corresponding audio recordings at a sampling rate of 10,000 Hz were also used in the current study to identify, by listening and spectrogram reading, the central frame position of 46 vowels in the 10 sentences. This subset of data frames is used in the analysis of vowels. A third subset contains 5 frames corresponding to the vowel [y]. Due to the fact that this articulatory model represents the vocal tract profiles with a relatively small number of parameters, there exist some errors in the parameters' values as pointed out in [11]. The accompanying computer programs of Maeda's model allow the reconstruction of the entire vocal

tract profile from the set of eight articulatory parameters. These programs can also transform the set of articulatory parameters into a set of uniform area sections. From this specification of the area function the length of the vocal tract is computed in this study for each of the 519 frames.

2.2. Analysis

Two types of statistical analyses are performed in this study. The first focuses on the distribution of the VTL parameter during the production of the 10 sentences and evaluates its mean, standard deviation, and range. The second is a regression analysis and calculates correlations between the VTL, lip X, and larynx height parameters, on one side, and the eight articulatory parameters, on the other side. The two types of analyses are performed on three sets of frames: first on all the 519 frames, second on the 46 frames corresponding to the selected vowels, and third on the 5 frames corresponding to vowel [y].

3. Results

The distribution of the VTL parameter over the 519 frames of the 10 sentences is presented in Figure 1. The mean and standard deviation of the VTL distribution is 15.05 cm and 0.53 cm, respectively. The VTL parameter covers a range between 14.00 cm and 17.29 cm, representing approx. 22% of the mean value. These statistics are obtained across all the phonemes in the 10 sentences. The 519 frames also include silence segments before and after sentences which sometime display variations in the positions of the lips and larynx height.

Linear regression analyses over the 519 frames show significant correlations between the VTL parameter and some other articulatory parameters. Figure 2 displays scattered plots between the VTL parameter and the eight articulatory parameters of Maeda's model. Each subplot also shows the corresponding correlation coefficient. These correlation coefficients are also presented in the second column of Table 1. It can be seen that the VTL is to some extent positively correlated with the tongue dorsum (0.63) and lip X (0.72) and inversely correlated with the larynx height (-0.63). To a less extent, the VTL is positively correlated with the tongue tip X (0.40) and negatively correlated with the lip Y (-0.39). As suggested in [2] the lip protrusion and larynx height appear to be major contributors to the length of the vocal tract. Surprisingly, these results show that another articulatory parameter has a significant contribution to VTL: the tongue dorsum parameter. This can be explained by the curvature shape of the vocal tract in which high values of the tongue dorsum increase the overall length of the vocal tract. Because the lip protrusion (lx) and larynx height (lh) can be controlled by the speakers relatively independently, this study analyzed their correlations with the other articulatory parameters. The results are presented in columns 3 and 4 of Table 1. A significant positive correlation was found between lip X and tongue dorsum (0.34), whereas a negative correlation was found between lip X and lip Y (-0.43). Larynx height is best correlated positively with tongue body (0.33) and negatively with tongue tip X (-0.33). All these correlations are all the 519 frames and thus are obtained across both vowels and consonants. A multiple linear regression analysis leads to the linear composition of the VTL described by Equation 1. The model accounts for about 95% of total VTL variability and the main contributions to the VTL parameter are from tongue dorsum, lip X, and larynx height.

$$VTL = 0.09\,j + 0.01b + 0.24d + 0.09tx + \\ + 0.02ty + 0.18lx - 0.05ly - 0.24lh + 15.05 \quad (1)$$

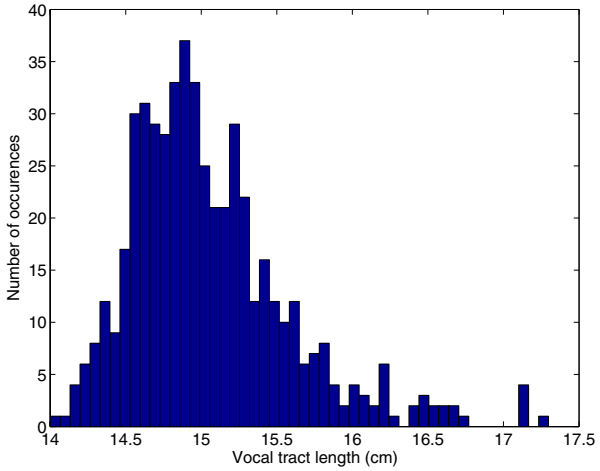

Figure 1: *VTL distribution over the 519 frames.*

Table 1. *Correlations between pairs of articulatory parameters in first row and first column: jaw (j), tongue body (b), tongue dorsum (d), tongue tip X (tx), tongue tip Y (ty), lip X (lx), lip Y (ly), larynx height (lh), and vocal tract length (VTL). Results based on all 519 frames.*

	VTL	lx	lh
j	0.18	0.20	0.12
b	-0.07	0.12	0.33
d	0.63	0.34	-0.10
tx	0.40	0.21	-0.33
ty	-0.06	-0.08	0.17
lx	0.72	1.00	-0.26
ly	-0.39	-0.43	0.24
lh	-0.63	-0.26	1.00

The above analyses were repeated for the smaller data set containing the central frames of 46 vowels. The VTL distribution over the 46 vowels has similar statistics: mean 15.04 cm, standard deviation 0.56 cm and range between 14.23 cm and 17.15 cm, representing approx. 19% of the mean VTL value. The corresponding correlation results are presented in Table 2. In general, the significant correlations obtained for the vowel frames are slightly higher than the corresponding ones for all frames. One might argue that these correlations, like in case of all frames, are averaged over various vowels, respectively phonemes, and thus they do not reflect phoneme specific phenomena. This is perfectly true, but it is interesting to see if significant correlations exist in running speech over all phonemes or over only vowels. The results just presented show that significant correlations among the articulators exist in such cases.

The next analysis is done on individual vowels. Four vowels were selected for analysis [a], [i], [u], and [y]. Due to space limitations, only the results for [y] are detailed in this paper. For this vowel the mean VTL is 15.07 cm, the standard deviation is 0.28 cm, and the range is from 14.77 to 15.50. The corresponding correlations, presented in Table 3, are obtained from only 5 occurrences of vowel [y].

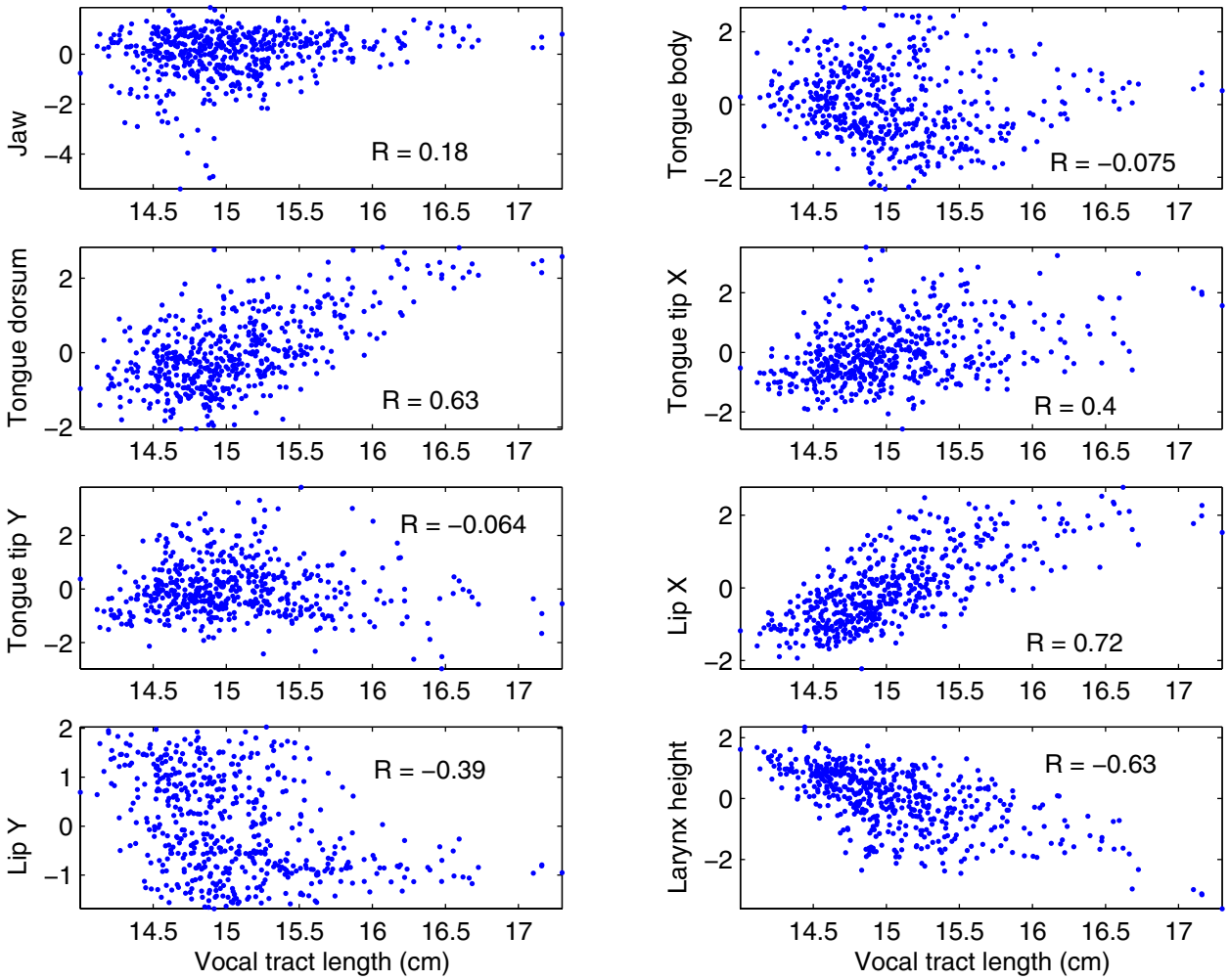

Figure 2: *Scattered plots for the 519 data frames between the VTL parameter and the eight articulatory parameters of Maeda's model. Each subplot displays the corresponding correlation coefficient R.*

Table 2. *Correlations as presented in Table 1 but based on 46 vowel frames.*

	VTL	Lx	lh
j	0.11	0.14	0.18
b	0.02	0.20	0.29
d	0.74	0.48	-0.25
tx	0.47	0.36	-0.44
ty	-0.16	-0.04	0.36
lx	0.77	1.00	-0.28
ly	-0.36	-0.43	0.04
lh	-0.65	-0.28	1.00

Table 3. *Correlations as presented in Table 1 but based on 5 frames of vowel [y].*

	VTL	lx	Lh
j	0.14	0.13	0.77
b	-0.02	0.42	0.63
d	0.55	-0.24	-0.17
tx	0.44	-0.06	-0.55
ty	-0.20	-0.32	0.37
lx	0.57	1.00	0.42
ly	-0.58	-0.87	-0.42
lh	0.27	0.42	1.00

Due to this fact, these statistics may not be very accurate in reflecting the real coarticulatory phenomena that exist for this vowel. The highest absolute correlations are observed between VTL and tongue dorsum (0.55), lip X (0.57), and lip Y (-0.58). It is interesting to note that for this small number of frames of [y] VTL is positively correlated with larynx height (0.27), unlike in the previous two cases. This could possibly be explained by the unexpected positive correlation between larynx height and lip X (0.42). If such positive correlation exists for the naturally produced vowel [y], this would support the suggestions made in [3], that larynx lowering may accompany insufficient lip protrusion, and would contradict the conclusions of [4]. However, both previous studies examined constrained speech whereas the current study examines natural speech. Another interesting observation for vowel [y] is the relatively strong negative correlation between lip X and lip Y (-0.87). The larynx height is positively correlated with jaw (0.77) and tongue body (0.63) and negatively correlated with the lip Y (-0.42). The correlation

between lip X and larynx height is negligible for vowel [u] (-0.03) and vowel [i] (-0.04), negative for vowel [a] (-0.46), and positive for vowel [y] (0.42). The correlation between VTL and lip Y is negative for [a] (-0.42), [i] (-0.52), and [y] (-0.58), and positive for [u] (0.62).

4. Discussion

Unlike other studies that examined the variability of articulators that contribute to the total length of the vocal tract [2-4], this study analyzes statistically the VTL variability and various correlations among articulators during production of natural speech by a French female speaker. The scattered plots obtained show that there are three main articulators significantly correlated with the VTL parameter: lip protrusion, larynx height, and tongue dorsum.

For vowels the mean VTL of the French female speaker examined in this study is 15.04 cm and the range is 2.92 cm, representing approx. 19% of the mean VTL value. For comparison, the mean VTL of a male speaker producing six Russian vowels is 17.83 cm and the range is 3.0 cm, representing approx. 17% of the mean VTL value [1]. These results show that the range of variation of the VTL parameter across the two speakers, in spite of having different genders, is approximately the same. However, more speakers are necessary for the analysis in order to generalize these conclusions.

All regression analyses were computed with *alpha*=0.01, corresponding to 99% confidence. The significant correlations showed high *F* and low *p* statistics (e.g. for VTL-lh in Table 1 *R*=-0.63, *F*=334, and *p*=0). The currently presented results show that significant correlations among various articulatory parameters exist in all the three types of data used (e.g. negative correlations between lip Y, on one side, and VTL and lip X on the other side. See Table 1, 2 and 3). Other coarticulatory phenomena can only be observed in specific phonemes (e.g. the correlation between lip X and larynx height is -0.46 for [a], -0.04 for [i], -0.03 for [u], and 0.42 for [y]; the correlation between VTL and lip Y is -0.42 for [a], -0.52 for [i], 0.62 for [u], and -0.58 for [y]). This shows that articulators have a relatively high potential of acting independently of each other, depending upon the phonological content.

The present study shows that a significant contribution to the overall length of the vocal tract is made by an articulatory parameter which was mainly overlooked by previous studies: the tongue dorsum. This parameter is contributing significantly to both shape and length of the vocal tract. The latter contribution is only possible due to the special curvature shape of the vocal tract. If the human vocal tract were straight, only the lip X and larynx height could change the VTL value during speech production.

In [8] it was suggested that speakers may vary the length of the vocal tract in addition to its shape in order to achieve desired acoustic targets. This appealing idea was also restated in [2] but a definite proof of this hypothesis is not provided yet. The significant variability of the lip X and larynx height parameters, as shown in Figure 2, appears to be indeed the result of a control strategy of speakers to achieve desired acoustic targets. Without the help of the lips and larynx height maneuvers, these acoustic targets would necessitate a larger displacement of the tongue parameters, as suggested in [8], which would lead to longer time and greater energy in achieving various phonological constructions. Thus, lip protrusion and larynx height movements appear to be articulatory strategies that help to optimize the speech production in terms of speed and energy while increasing the phonetic contrast. More work in supporting this appealing hypothesis is currently underway.

5. Conclusions

This paper presented a quantitative analysis of the variability of the length of the vocal tract during the production of natural speech by a single female French speaker. The data used in this study are limited to roughly about 10 seconds of speech. VTL varies significantly within a range of about 19% of the mean length in vowels and 22% in all 519 frames. This variability cannot simply be attributed to random causes and appears to be closely correlated with the constraints in achieving the desired acoustic targets of phonemes. Three articulatory parameters appear to contribute significantly to the instantaneous length of the vocal tract: lip X, tongue dorsum, and larynx height.

This study supports the conclusions in [7] that the interpretation of larynx height movements should be embedded in speaker-specific analysis of other articulators.

More data are necessary in order to draw definite conclusions on the variability of the vocal tract length and of related articulators, within speakers and across speakers.

6. Acknowledgements

This work was supported in part by the National Science Foundation under the ITR grant IIS-04-27413.

7. References

[1] Fant, G., Acoustic Theory of Speech Production. The Hague, The Netherlands: Mouton, 1960.

[2] Perkell, J. S., Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study. MIT, Cambridge, MA, 1969.

[3] Riordan, C. J., "Control of vocal-tract length in speech," J. Acoust. Soc. Amer., Vol. 62, pp. 998-1002, 1977.

[4] Tuller, B., and Fitch, H. L., "Preservation of vocal tract length in speech: A negative finding," J. Acoust. Soc. Amer., Vol. 67(3), pp. 1068-1071, 1980.

[5] Wood, S., "The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels," J. Acoust. Soc. Amer., Vol. 80(2), pp. 391-401, 1986.

[6] Sundberg, J., and Nordström, P.-E., "Raised and lowered larynx – the effect on vowel formant frequencies," STL-QPRS 2-3/1976, pp. 33-39. Speech Transmission Laboratory, Royal Institute of Technology, Stockholm.

[7] Hoole, P., and Kroos, C., "Control of larynx height in vowel production," 5th International Conference on Spoken Language Processing, Sydney, Australia, 1998.

[8] Mermelstein, P., "Determination of the vocal-tract shape from measured formant frequencies," J. Acoust. Soc. Amer., Vol. 41(5), pp. 1283-1294, 1967.

[9] Bothorel, A., Simon, P., Wioland, F., and Zerling, J. P., Cinéradiographie des Voyelles et Consonnes du Français, Travaux de l'Institut de Phonétique de Strasbourg, 1986.

[10] Maeda, S., "Une modèle articulatoire de la tongue avec des composantes linéaries," Actes 10émes Journées d'Etude sur la Parole, Grenoble, GALF, pp. 152-164, 1979.

[11] Maeda, S., "Compensatory articulation in speech: analysis of x-ray data with an articulatory model," Eurospeech'89, Paris, France, pp. 2441-2444, 1989.