Vowel and prosodic factor dependent variations of vocal-tract length

Shinji Maeda [1], *Yves Laprie* [2]

[1] CNRS LTCI; TELECOM ParisTech, Paris, France
[2] LORIA CNRS UMR7503, Villers-lès-Nabcy, France

shinji.maeda@mines-telecom.fr, Yves.Laprie@loria.fr

10.21437/Interspeech.2013-709

Abstract

We have measured total vocal-tract (VT) length, lip-tube length and glottal height during vowels on X-ray film data of short French utterances [1]. VT midpoints are determined by progressively fitting circles along the VT-length from the glottis to lip opening. The VT-length is obtained by summing up the distance between adjacent midpoints. Lip-tube length is defined as the distance between the incisors and the lip opening along the midline. Results show that the range of VT-length variation is 3.2cm with the average VT-length of 16.4cm. The cause of this large range appears to be the combination of the vowel dependent VT-length and prosodic position that influences on the glottal height. For example, during a vowel at sentence final position, glottis goes down with falling intonation or up with rising intonation corresponding to, respectively, VT lengthening or shortening. The lip-tube lengths are little affected by prosodic position and exhibit a clear vowel dependency. The prosodic influences manifested on the glottal height are not compensated but rather expanded in the VT-length, yet maintaining the characteristic vowel-dependency. This suggests an underlying mechanism to maintain a uniform stretching/compression of VT-length, which tends to hold the phonetic identity of a vowel under large VT-length changes.

Index Terms: speech production, vocal-tract length

1. Introduction

Perkell has suggested the preservation of a constant VT length control, observing articulators' position on the X-ray data [2]. Riordan has provided a support for Perkell's assertion from an upper-lip block experiment [3]. The blocked lip, presumably short, is compensated by an extra laryngeal lowering in labialized vowels /y/ and /u/. Tuller and Fitch however did not find any significant VT length preservation in the observation of lip protrusion and of laryngeal height for the vowel /u/ [4]. Moreover, Wood has claimed that a shortening of lip tube cannot be acoustically compensated by the extra lowering of the larynx, except for the first formant frequency [5].

We feel that the discussion about the VT length of these authors lack the measurement of the VT length. To estimate the VT length, we have to refer to articles dealing with the derivation of the VT area functions, e.g. [6, 7, 8, 9, 10] where the main concern is shapes. For example, six Russian vowels exhibit large vowel dependent VT length where the /u/ is longest (19.5cm) and /a/ shortest (16.5cm) with the average length of 17.8cm over the six vowels.

In this study, we investigate the variation of the VT length. For this purpose, we shall formulate a method for the determination of the VT midline to measure the length on the VT midsagittal contours derived from the X-ray film. Moreover, we analyze VT-length variations during vowels that occur in sentences so that measured VT-length variations would include the prosodic effects which have been excluded in the previous literature.

2. Corpus and Data

The material for our investigation is digitized VT contours hand-traced on the projected X-ray film images (50 frames/s). It consists of ten short French sentences uttered by a female speaker [1]. Table 1 lists the sentences and the number of X-ray frames in each sentence in which the VT is not occluded, as vowels and non-occlusive consonants, so that total VT length from the glottis to the lip opening can be measured. Our investigation here is concerned with vowel segments.

Table 1: Sentences and the number of frames in which the VT is not completely closed

ID	Sentences	# of frames
pb01	[ma ʃmiz ɛ ʁusi]	48
pb02	[vwa la dø buʒi]	33
pb03	[dɔn œ̃ (p) ti ku]	24
pb08	[yn ʁepɔ̃s ɑ̃mbigy]	38
pb09	[lwi pɑ̃ sa sɑ]	53
pb15	[mɛ te bozabi]	38
pb17	[yn pata ʃu]	36
pb18	[pʁɛt lwi sɛʒ eky]	45
pb24	[ʃəvalje dy ge]	32
pb28	[il fym sɔ̃ tabɑ]	39
total		386

In each frame, the VT profile is represented by the ordered series of contours corresponding to the upper lip, incisor, hard palate, velum, rear pharyngeal wall, and rear laryngeal wall ending at the posterior edge of the glottis, which constitute the outer VT outline. Similarly, contours of the lower lip, incisor, tongue, epiglottis, and anterior laryngeal wall ending at the anterior end of the glottis constitute the inner VT outline. Points representing individual contours are ordered from the glottal end to the lip opening and a spline interpolation assures the maximum spacing between adjacent points to be less than .3mm. These refinements are necessary for the proper operation of the circle fitting method described in the next section. All the VT profiles are registered in relative to the head position and size.

3. Determination of VT midline

The determination of midline appears to belong to art than to science. It is not so evident to formally define the midline of VT that has an arbitrary geometry. Our strategy here is to define the geometrical midpoint for a simple tube such as a conical tube, and then to determine midpoints along the VT by applying this definition.

Figure 1 illustrates our definition of the midpoint assuming a tube having conical shape. The two thick lines, OM and ON, depict the walls. The circle with its center P just touches walls at point A and B. Geometrical operation shows that the center point P and midpoint C of the line AB are right on the axial line OQ, i.e. the midline of the conical tube.

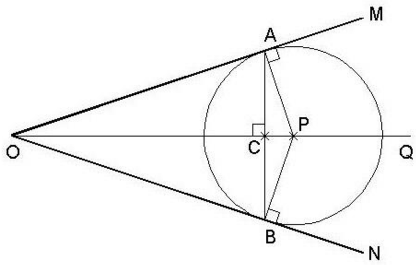

Figure 1: Definition of midpoint C. The circle P is tangential to the walls, OM and ON, of the conical tube, respectively, at point A and at B. The midpoint of the line AB (also the circle center P) is on the axis OQ, i.e. the geometrical midline of the conical tube. The line AB (actually a surface) can be regarded as a theoretical plane-wave front of propagating sound inside the tube.

It may be noteworthy to point out that the line AB is considered as a plane-wave front, if the plane-wave propagation is assumed. Although the plane wave propagation inside VT is a gloss approximation, the definition of midpoints with the line AB is consistent with the concept of area function to calculate VT acoustics.

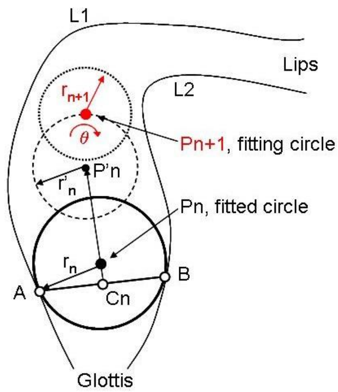

Figure 2: Iterative procedure for the determination of fitting circles through the glottis to the lips. The sequence of midpoints C_n constitutes the VT midline.

Figure 2 illustrates how the VT midpoints are determined by fitting circles. In the conical tube as described above, the circle can be analytically derived, but this is not the case in the real situation. The outer (L1) and inner outline (L2) are presented as sequences of points where their position is specified by xy-coordinates. Moreover, shapes of these contours are arbitrary. The condition "just touch", therefore, is defined such that the circle contains a single point on L1 and another single point on L2. Actually however, this single-point condition must be loosened to a few points or less to ensure the stability of fitting process.

The initial starting midpoint is defined, exceptionally, as the center point between L1 and L2 glottal end. The radius of this initial circle is the half of the glottal length for the sake of convenience. At the n-th step, let us assume, as seen in Figure 2, that the circle P_n just fits at A on L1 and B on L2. Its center is at P_n and radius r_n. The determined n-th midpoint is C_n. The next midpoint C_{n+1} will be determined by a set of heuristic rules.

The first two rules determine a circle P_n' (indicated by the dashed circle) along which the center of next circle P_{n+1} can move. The position of a circle P_n' is determined on the extended line along the line C_nP_n. The distance between P_n and P_n' is specified in function of a weighted r_n, which controls a progression rate as,

R1: distance $\overline{P_nP_n'} = k_1 r_n$

and its radius that should be smaller than r_n by

R2: radius $r_n' = k_2 r_n$,

where k_1 and k_2, and also $k_3 \sim k_6$ below, are arbitrary constant whose values must be empirically determined.

The position of the next fitting circle P_{n+1}, shown in the dotted circle, is iteratively adjusted its position on the dashed circle P_n' and its radius r_{n+1} to achieve the fitting by applying basically following a top-down series of five rules. The variables m_1 and m_2 are the number of contour points located inside the fitting circle P_{n+1}, respective, on L1 and L2.

R3: If m_1=1 and m_2=1, the fitting is done and go to R1.

R4: If m_1>0 and m_2>0, reduce radius by $r_{n+1} \rightarrow r_{n+1} - k_3$ and go to R3 (with a lower limit on r_{n+1}).

R5: If m_1=0 and m_2=0, increase radius by rewriting as $r_{n+1} \rightarrow r_{n+1} + k_4$ and go to R3 (with a upper limit).

R6: If m_2>m_1, move P_{n+1} toward L1 by $\theta \rightarrow \theta - k_5$ and go to R3.

R7: If m_2<m_1, move P_{n+1} toward L2 by $\theta \rightarrow \theta + k_6$ and go to R3.

The midpoint detection process stops as the radius r_{n+1} reaches to the lower limit such as .5mm in R4 and also to the upper limit such as 4cm in R5. The lower limit is to detect the presence of a strong constriction or closure, whereas the upper limit detects the exit of the fitting circle from the lip opening as seen in Figure 3a. We must admit the heuristic nature of the process is a major weakness of our rule-based method. In the detection on the 386 VT contours, it resulted in 29 failures (7%), which is acceptable for this analytical study nonetheless.

An example of fitted circles and the determined VT midline are shown, respectively, in Figure 3a and 3b. Three interesting features of the circle fitting method are visible in Figure 3a: 1) The circles follow the principal tract gracefully ignoring the side branches such as the piriform fossa next to the larynx and the sublingual cavity between the tongue apex and the lower incisor. 2) The density of the circles varies in function of the tract width due to rule R1, which results in a finer specification of VT geometry in narrow zones than in wider zones. 3) The end of VT is automatic, because plane-wave front is blocked at the lip edges even after fitted circles go out of the VT. Although the acoustic validity of such a VT termination might be questioned, the automatic procedure assures its consistent determination.

The determined wave fronts and the VT midline are shown in Figure 3b. The wave fronts are often not perpendicular to the VT midline. The cause of this discrepancy could be due to the locally asymmetry of the outer and inner VT outlines relative to the axis, to the varying density of points that defines outlines and to the non-optimality of the detection rules. Nevertheless, the determined VT midline appears to be relatively smooth without any zigzag that would artificially lengthen the VT length. We feel therefore that it is adequate for the calculation of the VT length in this study.

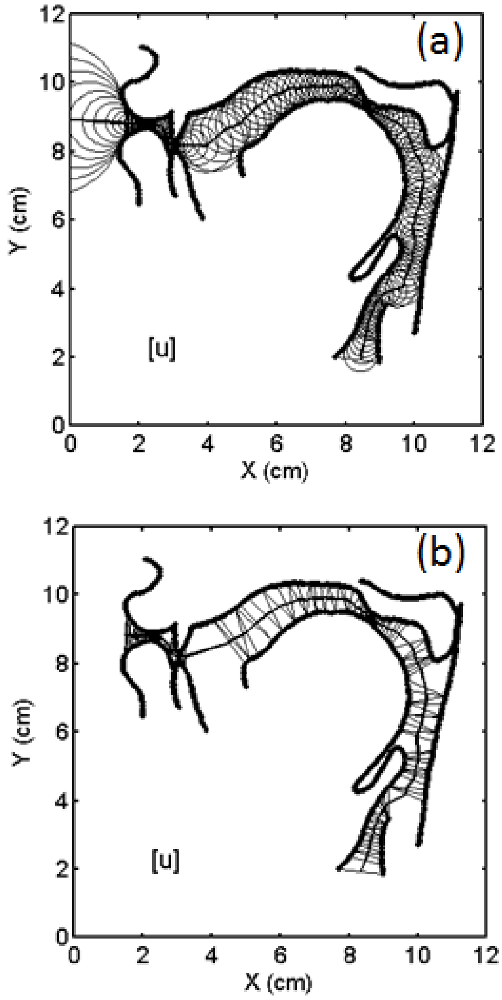

Figure 3: Midsagittal VT contours of the vowel [u] selected from the utterance pb01 in Table1, in (a) fitted circles and the line connecting the center of circles and in (b) a series of plane-wave front and the determined VT midline.

4. Results: VT length variations

In order to understand better the nature of VT length variations, we consider two additional length-related variables, lip-tube length and the glottal height measured by the y-coordinate value of the center of the glottis. As mentioned earlier, the lip-tube length is the distance between the incisors and the lip-opening edge along the VT midline.

Figure 4 shows overlays of raw frame-by-frame variations of these three variables during oral vowels extracted along the 10 French sentences. Considering a small number of samples, we felt reasonable to group [e] and [ɛ], [a] and [ɑ], and [ɔ] and [o]. We have noticed that VT length is often longer when the same vowel is located in the sentence initial or final position than in a middle position. The dotted lines in Figure 4 indicate the former cases and the solid lines the latter cases. Note that the y-axis of the glottal height in the bottom panel is reversed. In this way, glottal lowering becomes comparable to the VT and lip-tube lengthening and the glottal rising to the shortening, as in the other two panels.

5. Discussions

The vowel dependency and prosodic effects are evident in Figure 4. Let us start to discuss about the prosodic effects. The effect of the prosodic position is evident, especially, in the glottal height. The dotted curves systematically occupy the upper limit of the dispersion for the all vowels. A clear separation of dotted over solid curves are seen for [i], [y], and [u]. In these vowels, the rising curves, indicating the VT lengthening, appear to be the direct consequence of the glottal lowering in final position with concomitant F0 lowering. The two dotted curves marked by the arrows, one in [e] from the utterance pb24 and the other in [ɑ] from pb28, manifest a falling pattern, which corresponds to glottal rising and thus to a VT shortening. These vowels are in sentence final position. Interestingly the subject ended these two sentences with rising F0, presumably due to the listing-effect, during the acquisition of the X-ray film data.

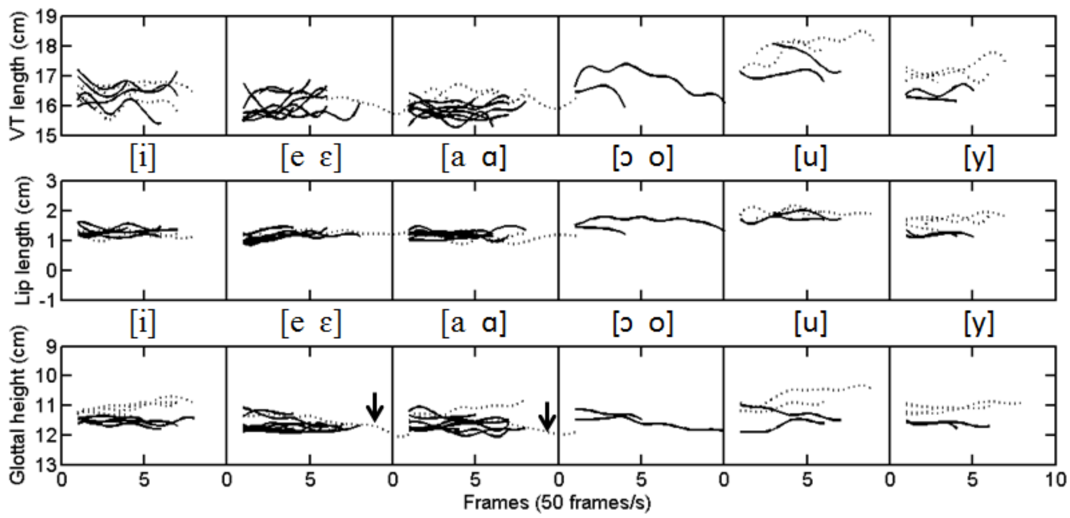

Figure 4: From top to bottom, onset aligned frame-to-frame variations of VT length, of lip-tube length, and of reversed glottal height in vowel segments extracted from sentences listed in Table 1. The solid curves are for vowels in sentence middle position and the dotted ones for those in sentence initial or final position. Similar vowels such as [a] end[ɑ] are grouped.

The prosodic position doesn't change the lip-tube length, except in the case of [y] and to a lesser extent [u]. Only labialized vowels, therefore, can exhibit prosodic effect upon all three length variables. In non-labialized vowels, the prosodic effect appears to be minimal. It is interesting to observe, regardless of absence or presence of the prosodic effect, that the magnitude of the total VT-length variations seems to expand more than the sum of individual variations of the lips and larynx. This expansion suggests that the tongue position is another determinant for VT length. Indeed, Perkell has pointed out that VT length is controlled through an interaction of the larynx height, degree of opening of the mandible, and protrusion of the lips (p.65 in [2]). We prefer here to replace the mandible by the tongue that directly affects upon the VT shapes. The expansion of VT-length variations of the labialized vowels, at least, suggest that all the VT parts, the lip-tube, oral and pharyngeal/laryngeal cavity, lengthen in a not compensatory but in a synergetic manner. It may be stated then that our French subject tends to control uniform length changes throughout VT, presumably, to preserve the phonetic value of labialized vowels against the large VT-length variations.

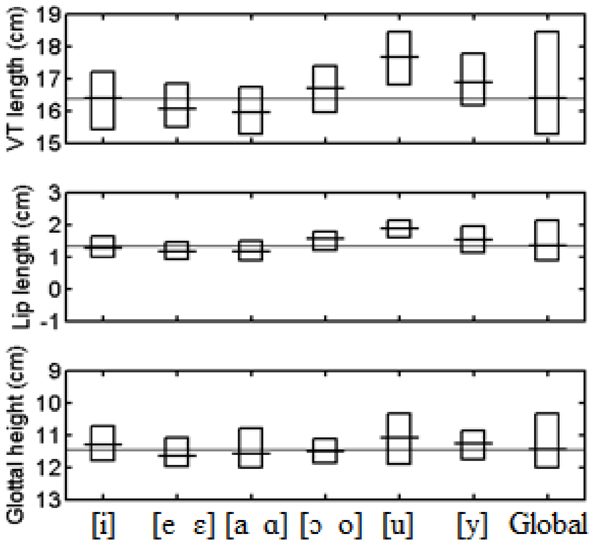

Figure 5: Rectangles indicate min/max ranges of total VT length, lip-tube length and reversed glottal height averaged over individual vowels. The short horizontal lines across rectangle show the vowel-averaged values. The rectangles labeled as "Global" are for all the vowels together. The thin horizontal line across each panel indicates global average value as a reference.

The other visible attribute of those data curves in Figure 4 is the vowel-dependency that is better illustrated in Figure 5. The shapes of vowel-dependent pattern of VT length, of lip-tube length and even of glottal height, which is strongly affected by vowel position, are quite similar to each other. VT length and lip-tube length of the non-labialized vowels, mostly front ones, are equal to or below the global average, whereas the labialized vowels, back ones, are above the global average. This kind of vowel-dependent pattern is seen in the VT area function data, e.g., p.115 in [6]. The similarity of vowel dependency patterns of VT length, lip-tube length and glottal height could be a part of articulatory specifications for the individual vowels. If this is the case, the vowel average values shown in Figure 5 may be regarded as the intrinsic VT length of vowels.

In order to explain the VT length difference in labialized and non-labialized groups, we need again to refer to the tongue position. The tongue position appears to be an important determinant for VT length. In our rough estimation, the tongue-position is dominant in VT-length change from [a ɑ] to [i], and the tongue position and lip-tube length exert about equal contribution of in from [a ɑ] to [u].

Now, let us discuss qualitatively how could change the length of the oral cavity (front) cavity, and pharyngeal (back) cavity in function of the tongue position. By the way, we assume that glottal-height changes affect mainly on the length of the pharyngeal tube and not on that of laryngeal tube.

Let us consider the situation where the tongue position is changed from a neutral position, i.e., schwa position, to extreme vowels, [i], [a], and [u]. For the [i], tongue moves toward a high/front position; the front cavity, therefore, should be shortened and the back cavity lengthened, in a compensatory manner. For the [a], the compensatory length changes of these two cavities should occur, but in this case the lengthening of front the cavity is compensated by the shortening of the back cavity. In non-labialized vowels therefore, the length variations due to changes in tongue position are relatively small because of these compensations. Situation is different for [u]: Because of its high/back tongue position, both front and back cavity would undergo concomitant lengthening, resulting in a synergetic VT elongation.

From the above discussion, it may be reasonable to state that the characteristic vowel-dependent VT-length variation comes from the vowel-specific articulatory position of the tongue itself, which is enhanced by the concomitant length variation of the lips.

Finally the solitary behavior of the vowel [y] deserves our attention. As seen in Figure 5, the average length of [y] is longer than that of vowels [ɔ and o], even though the tongue position of [y] is closer to that of [i]. In fact, its long VT length is primarily due to an extra lip protrusion in the sentence final position, as indicated by the two dotted lines in Figure 4. The averaging calculations of VT length excluding these dotted lines have resulted in lengths comparable to that of [i]. It may be noted that the French vowel [y] is produced with lip protrusion without rounding. This is to lengthen the front cavity consisting of the oral cavity and lip tube. Presumably, this extra lip protrusion goes with the prosodic glottal lowering at the final position leading a uniform VT lengthening described before.

6. Concluding remarks

We have demonstrated that the vowel identity and prosodic position determine VT length during vowels. We have postulated underlying mechanisms operating in VT length variations in a qualitative manner. Since our investigation is based on a small body of data, the hypothetical mechanisms must be confirmed or infirmed by a quantitative examination on a larger body of new data, e.g., from MRI movie.

7. Acknowledgements

This work is supported in part by ARTIS project, EMER-001-01, financed by the French National Research Agency (ANR).

8. References

[1] Bothorel, A., Simon, P., Wioland, F., and Zerling, J.-P. “Cinéradiographie des voyelles et consonnes du français, Travaux de l’Institute de Phonétique de Strasbourg, France, 1986.

[2] Perkell, J. S., “Physiology of speech production: Results and implications of a quantitative cineradiographic study”, The MIT Press, 1969.

[3] Riordan, C. J., “Control of vocal-tract length in speech”, J. Acoust. Soc. Am., 62(4), 998-1002, 1977.

[4] Tuller, B., and Fitch, H. L., “Preservation of vocal-tact length in speech: A negative finding”, J. Acoust. Soc. Am., 67(30), 1066-1071, 1980.

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

[6] Fant, G., “Acoustic theory of speech production”, Mouton, The Hague, (2nd Edi.), 1970.

[7] Bear, T., Gore, J. C., Gracco, L. C., and Nye, P. W., “Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels”, J. Acoust. Soc. Am., 90, 799-828, 1991.

[8] Perrier, P., Boë,. L.-J., and Sock, R., “Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal-tract cast”, J. of Speech and Hearing Research, 35, 53-67, 1992.

[9] Takemoto, H., Honda, K., Masaki, S., Shimada, Y., and Fujimoto, I., “Measurement of temporal changes in vocal tract area function from 3D cine-MRI data,” J. Acoust. Soc. Am. 119, 1037–1049, 2006.

[10] Clément, P., Hans, S., Hartl, D. M. Maeda S., Vaissieèe, J., and Brasnu, D. "Vocal tract area function for vowels using three-dimensional magnetic resonance imaging: A preliminary study", J. of Voice, 21(5), 522-530, 2007.