Velocity Differences Between Velum Raising and Lowering Movements

Peter Birkholz(✉) and Christian Kleiner

Institute of Acoustics and Speech Communication, Technische Universität Dresden, Dresden, Germany
{peter.birkholz,christian.kleiner}@tu-dresden.de

Abstract. This study investigated the intrinsic velocities of raising and lowering movements of the velum that are related to its biomechanical structure and aerodynamic conditions. To this end, five subjects produced cyclic transitions between nasals and fricatives as in /s-n-s-n-s-.../ with flat intonation and at two specific speaking rates to minimize contextual and prosodic effects. The velar movements were inferred from the movements of the lateral pharyngeal wall in ultrasound image sequences, which are strongly correlated. The results indicate that velum raising was significantly faster than velum lowering for the two male subjects (24%–49% faster, depending on the subject and speaking rate), but not for the three female subjects. Possible biomechanical and aerodynamic reasons for the observed velocity differences are discussed. The results can inform the interpretation of kinematic data of velar movements with regard to underlying neural control, and improve movement models for articulatory speech synthesis.

Keywords: Speech production · Velar movement · Articulator velocities

1 Introduction

The velocity of an articulator when it approaches the target for a certain speech sound depends on its biomechanical properties, on the active control by the nervous system, and on aerodynamic conditions [11]. The individual contributions of these factors to an observed articulatory movement are of great interest in speech production research, but they are usually hard to disentangle. When one articulator moves *on average* slower or faster than another one across different phonetic contexts and aerodynamic conditions, this difference is likely due to their biomechanical differences. For example, the movement of the tongue towards and away from a velar closure is generally slower than that of the lips during the formation or release of a bilabial closure [24], which is likely related to the different masses of the tongue and the lips.

Furthermore, there is evidence that the velocities of articulators also depend on the *direction* of movement. For example, it was observed that lip retraction is on average faster than lip protrusion [7], that tongue backing is on average slower

A. Karpov and R. Potapova (Eds.): SPECOM 2021, LNAI 12997, pp. 70–80, 2021.
https://doi.org/10.1007/978-3-030-87802-3_7

than tongue fronting [1,8,27], that vocal fold adduction is slower than vocal fold abduction [15], and that the maximum speed of pitch change is significantly higher for pitch lowering movements than for pitch raising movements [28].

For velar movements, it is not completely clear yet whether there is an intrinsic (i.e. biomechanical) difference between raising and lowering velocities. While studies on velar movements during the articulation of European Portuguese nasal vowels indicated that velum lowering takes on average longer than velum raising [18,21,25], studies with French nasals and nasalized vowels reported different observations: One study found that the durations of velum raising and lowering are essentially equal, independently from the nasal vowel, its phonetic context and the speaking style [2], while another study found indications that velum raising is faster than lowering [5]. Since the recording of velar movements is rather difficult, all of these studies were based on only 1–3 subjects. Furthermore, they were not specifically designed to explore velocity differences.

The goal of the present study was to gain more insight into the intrinsic direction-dependent velocity differences of the velum using a tailored experimental design and more subjects. The subjects produced alternating raising and lowering movements of the velum in phoneme sequences like /s-n-s-n-s-.../ with flat intonation and specific speaking rates. In this way the phonetic context and prosodic factors were precisely controlled, so that differences between the observed raising and lowering movements can be attributed solely to biomechanical and aerodynamic factors.

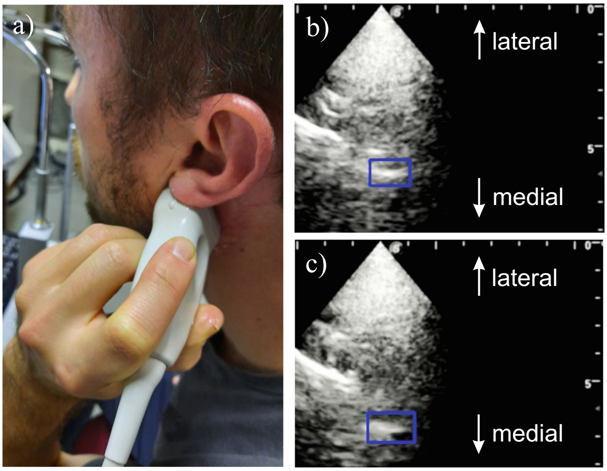

Fig. 1. a) Positioning of the ultrasound transducer for measuring lateral pharyngeal wall movements. b) Ultrasound image for a pharyngeal wall position that corresponds to a lowered velum. c) Ultrasound image for a pharyngeal wall position that corresponds to a raised velum. The blue boxes in b) and c) indicate the tracked region of interest.

As mentioned above, the measurement of velar movements is difficult because of the individual drawbacks of the existing techniques [16]. Real-time MRI of the vocal tract is rather expensive, and the subjects are in supine position [18]. X-ray cinematography [19] is normally not used in speech studies anymore because of its ionizing radiation. Other methods like fiberoptic endoscopy [3–5], electromagnetic articulography [25], or velum-specific measurement devices like the Velotrace [13] or the Nasograph [9] are invasive and unpleasant for the subjects. In the present study, we therefore used an indirect method to track velum movements based on ultrasound. It is based on the observation that the lateral pharyngeal walls (LPW) move medially when the velum raises, and laterally when the velum lowers [22]. LPW movements and velar movements are highly correlated, i.e. parallel in both time course and extent [3,4], and can be detected on ultrasound images when the transducer is placed below one ear of the subject as illustrated in Fig. 1a.

Learning about intrinsic direction-dependent velocities of articulators cannot only benefit the interpretation of kinematic speech data, but can also help to improve the realism and quality of articulatory speech synthesis [6,10,26]. For example, it was found that different intrinsic velocity components for different directions of tongue movement can explain and reproduce the observed loop patterns in tongue trajectories [8,27]. Modeling such more realistic tongue trajectories (as opposed to straight-line paths between the articulatory targets) was found to lead to more natural synthetic speech generated by articulatory synthesis [14,20]. A better model for the movement of other articulators like the velum might further improve the realism and quality of articulatory synthesis.

2 Method

2.1 Subjects and Corpus

Two male and three female German speaking subjects (29–43 years old) participated in the experiment. All subjects gave informed consent and reported no speech or hearing disorders. Each subject produced two times the 16 phoneme sequences listed in Table 1 while the speech audio signal and the velar movements were recorded.

Table 1. Phoneme sequences uttered by the subjects.

Index	Sequence	Rate	Index	Sequence	Rate
01	/s-n-s-.../	slow	09	/s-n-s-.../	fast
02	/s-m-s-.../	slow	10	/s-m-s-.../	fast
03	/f-n-f-.../	slow	11	/f-n-f-.../	fast
04	/f-m-f-.../	slow	12	/f-m-f-.../	fast
05	/n-s-n-.../	slow	13	/n-s-n-.../	fast
06	/m-s-m-.../	slow	14	/m-s-m-.../	fast
07	/n-f-n-.../	slow	15	/n-f-n-.../	fast
08	/m-f-m-.../	slow	16	/m-f-m-.../	fast

Each phoneme sequence consisted of at least five repetitions of a fricative-nasal or nasal-fricative pair. The fricatives and nasals with the alveolar, labiodental, and bilabial places of articulation were selected in such a way that the velum had no or minimal contact with the tongue back at any time for as unrestricted velar movement as possible. Furthermore, the movements of other articulators were small, leading to a low interference with the velar movement. For each pair of a fricative and a nasal there was one sequence that started with the nasal, and one sequence that started with the fricative. The phonemes in a sequence were produced with a flat intonation and at a specific rate as dictated by a metronome (without glottal stops between the phonemes). The slow and fast sequences were produced with 500 ms and 375 ms per phoneme, respectively. With this experimental design, contextual and prosodic factors were precisely controlled so that any potential differences between velar raising and lowering movements can be attributed to biomechanical or aerodynamic factors.

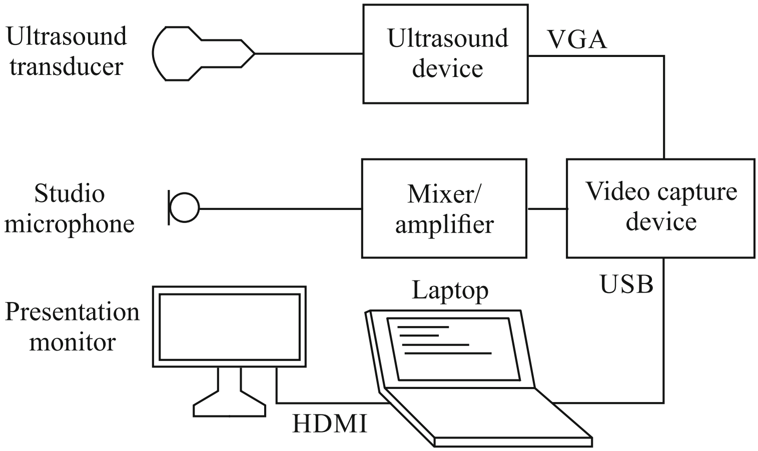

Fig. 2. The measurement setup used in the experiment.

2.2 Experimental Setup and Recording Procedure

Velar movements were measured in terms of lateral pharyngeal wall (LPW) movements as mentioned in the Introduction. This is possible because lateral-medial movements of the LPW are highly correlated with lowering-raising movements of the velum [3,4,23]. The experimental setup is shown in Fig. 2. B-mode scans of the LPW were captured with a medical ultrasound device (SonoScape S2 using a SonoScape 2P1 transducer) at 1.8 MHz with a rate of 50 frames per second. The speech audio signal was recorded with a studio microphone (M930 by Microtech Gefell) fed by a mixer (Behringer MX1602). A video capture device (USB3HDCAP by StarTech) digitized the audio signal and the ultrasound video frames, which were resampled at a rate of 60 frames per second (by doubling every 6th frame). The digital signals were sent via a USB connection to a laptop computer (MSI

GT72 with MS Windows 8.1) where they were recorded and saved with the video grabber software StreamCatcher.

The experiment was performed in a recording studio in individual sessions for each subject. The subjects were seated in a comfortable position with the head stabilized by a headrest. Before the actual recordings the subjects practiced the timed phoneme changes of the utterances. For the recordings, the ultrasound transducer was held by an assistant at a position below the left ear as illustrated in Fig. 2a. The orientation of the probe was carefully adjusted to obtain as clear a picture as possible of the LPW movement. Whether an orientation was suitable was tested with the utterances /n-s-n-s-.../ and /f-s-f-s-.../, where clearly visible movements were expected for the first utterance, and no movements for the second utterance (due to a constantly high velum position). After a suitable probe orientation was found, the utterances in Table 1 were displayed one by one on a presentation monitor and spoken by the subject in two rounds. The phoneme changes after 500 ms (slow) or 375 ms (fast) per phoneme were indicated by click sounds played over the laptop loudspeaker. In the case of ultrasound image artifacts or an incorrectly produced sequence, the utterance was immediately repeated.

2.3 Tracking of the Lateral Pharyngeal Wall Movement

Using the audio data as time reference, the captured ultrasound video files were segmented into the individual utterances using the software VirtualDub2 (http://www.virtualdub.org). Using a custom-made Python script, the LPW movement in each utterance was tracked across the ultrasound image sequence based on the region tracker [17], which is available as the class TrackerCSRT in the Open Source Computer Vision toolbox 4.5 (https://opencv.org). Since not only the position but also the appearance of the LPW in the ultrasound images often changed between the raised velum state and the lowered velum state, the initial frame and region of interest for the tracker had to be carefully selected for each sequence. In 23% of the recorded utterances the tracker failed to follow the LPW movements over at least three velar raising-lowering periods and the corresponding utterances were discarded[1].

Figures 1b and c show example video frames for the two velum states. The tracked region of interest is marked by the blue rectangles. The lower rectangle position in Fig. 1c indicates a more medial LPW position and hence corresponds to a raised velum. An example of the tracked horizontal and vertical positions of the LPW (center of the tracked region) in the ultrasound images over an entire utterance is shown in Figs. 3a and b. Because there was usually not only movement along the vertical axis but also along the horizontal axis, the main movement direction in the image plane was determined by means of a principal component analysis, and the 2D-trajectory was then projected on the first principal component to obtain a one-dimensional position signal (Fig. 3c).

[1] The difficulty of finding a good ultrasonic LPW reflection was also reported in [22].

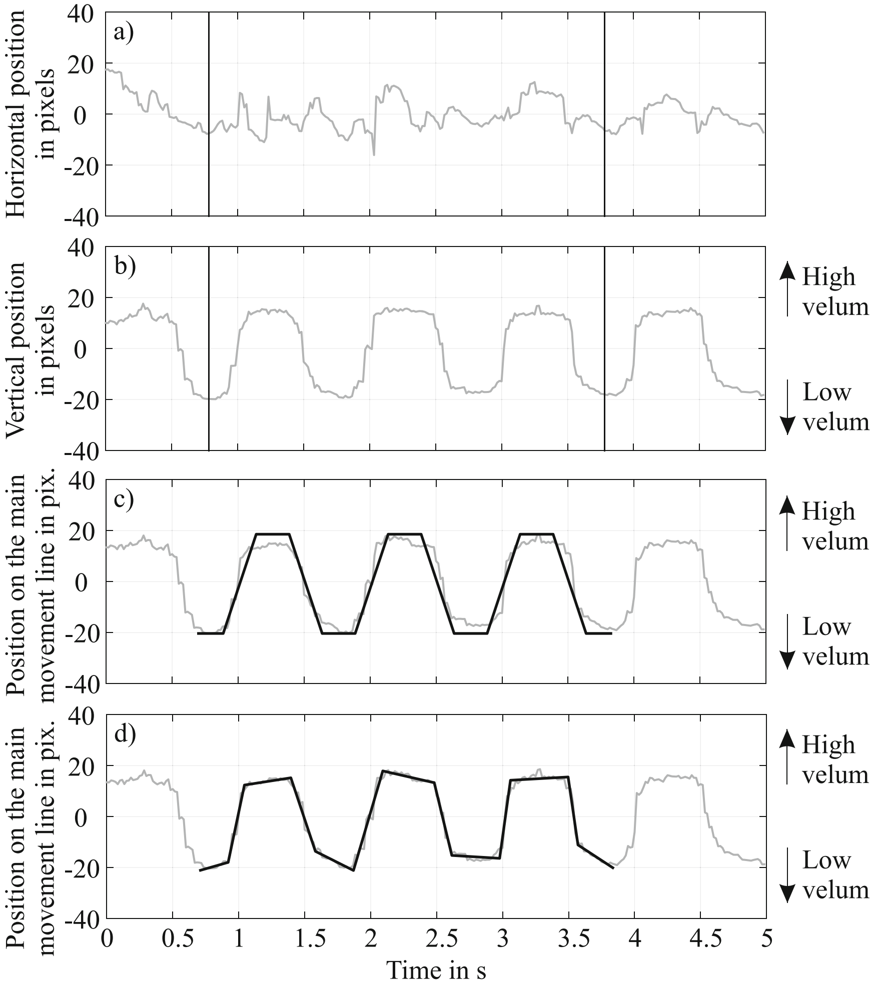

Fig. 3. a, b) Time functions of the horizontal and vertical positions of the lateral pharyngeal wall on the ultrasound images for the utterance /m-s-m-s-.../ (slow speaking rate) by speaker M2. c) Lateral pharyngeal wall position projected on the main movement line (first principal component of the x-y-trajectory; gray curve), and initial approximation by straight-line pieces (black curve). The main movement line was calculated for the trajectory between the vertical black lines in a) and b). d) Approximation of the gray curve in c) by line pieces (black curve) after optimization. Increasing position values in c) and d) correspond to a medial movement of the nasopharyngeal wall and hence to an increasing height of the velum.

2.4 Determination of Movement Velocities

In each LPW position signal three consecutive periods were approximated by a sequence of straight-line pieces. Each period was represented by four line pieces:

one for the raising edge, one for the quasi-stationary phase with the raised velum, one for the falling edge, and one for the quasi-stationary phase with the lowered velum. The line pieces were first prototypically initialized as illustrated by the black lines in Fig. 3c. The endpoint coordinates of the line pieces were then automatically optimized in such a way that the root-mean-square error between the original position signal and its piece-wise linear approximation was minimized (Fig. 3d). The optimization was performed with the function fminsearch in Matlab R2019b, which is a derivative-free simplex search method. The (absolute) slopes of the line pieces for the raising and falling edges (with the unit pixels/second) were taken as approximations of the raising and lowering velocities of the velum, respectively. This approach based on line pieces was more robust against noise and prevented ambiguities compared to other methods, like e.g. smoothing the curve and then taking the maxima and minima of the first derivative as approximations of the velocities.

In a few cases, problems of LPW tracking led to unnaturally short transition phases. These cases were detected by comparing the durations of the raising and falling edges with a threshold, and all edges with a shorter duration were discarded. Since the minimum duration of a complete alternating movement cycle of the velum is in the range of 200–300 ms [24], a conservative threshold of 60 ms was chosen here. Finally there were between 19–40 rising edges and 16–42 falling edges per subject and speaking rate available for statistical analysis.

Table 2. Mean velocities of the raising and lowering movements of the velum in pixels/second ($\overline{v}_{\text{raise}}$ and $\overline{v}_{\text{lower}}$, respectively) at the slow and fast speaking rates for all five subjects. The p values indicate whether or not the differences are statistically significant according to t-tests. Significant differences ($\alpha = 0.01$) are indicated by bold p values. The lowest row presents the results pooled over all five subjects.

Subject	Slow speaking rate			Fast speaking rate		
	$\overline{v}_{\text{raise}}$	$\overline{v}_{\text{lower}}$	p	$\overline{v}_{\text{raise}}$	$\overline{v}_{\text{lower}}$	p
M1	191	145	**0.010**	253	170	**0.002**
M2	278	210	**0.004**	287	232	**0.006**
F1	206	216	0.769	238	227	0.590
F2	260	269	0.747	318	277	0.302
F3	265	234	0.102	234	229	0.801
all	248	217	**0.007**	267	226	**0.000**

3 Results

Figure 4 shows the distributions of the raising and lowering velocities of the velum individually for all five subjects and both speaking rates, and pooled across all subjects for each speaking rate. Table 2 summarizes the mean values $\overline{v}_{\text{raise}}$ and $\overline{v}_{\text{lower}}$ of these distributions. Except for two cases (subjects F1 and F2 at the slow speaking rate) the mean raising velocity of the velum was higher

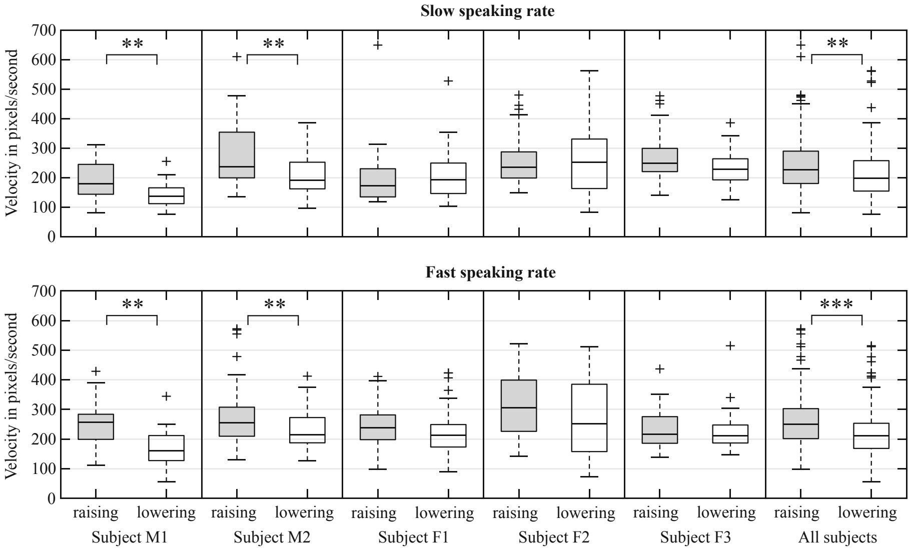

Fig. 4. Distributions of the raising and lowering velocities of the velum (inferred from LPW movement) at the slow and fast speaking rates for all five subjects individually, and pooled across all subjects. Significant differences between the raising and lowering velocities are indicated with ** ($p < 0.01$) or *** ($p < 0.001$).

than the lowering velocity. Based on two-sample two-sided t-tests, the differences were significant for the two male subjects M1 and M2 ($p < 0.01$), and when the data were pooled across all subjects ($p < 0.001$). For the two male subjects, raising movements were 24%–49% faster than lowering movements, depending on the subject and the speaking rate. Across all subjects, velum raising was on average 14% faster than velum lowering for the slow speaking rate, and 18% faster for the fast speaking rate. From the slow to the fast speaking rate, the velocities increased on average by 7.7% and 4.1% for velum raising and lowering, respectively.

4 Discussion and Conclusion

This study found that velum raising was faster than velum lowering for two out of five subjects, which confirms the findings of the studies [5,18,21,25], and that there was no significant difference for three subjects, which confirms the observations of [2]. In contrast to the previous studies, the present study included more subjects and was specifically designed to uncover potential direction-dependent velocity differences. Confounding factors like prosodic and contextual variations, or physical contact between the velum and the tongue were minimized. This leaves essentially biomechanical/muscular or aerodynamic reasons for the observed velocity differences.

With regard to biomechanics, it is generally accepted that velum raising (and retraction) is caused by the activity of the levator palatini [4]. However, the mechanism for velum lowering is not as clear. Some researchers suggest that velum lowering is implemented by the joint activity of the palatopharyngeus and palatoglossus [12]. In contrast, the reviewed EMG studies by Bell-Berti [4] do not provide support for the role of the palatopharyngeus as a velar depressor and had inhomogeneous conclusions for the role of the palatoglossus, i.e., palatoglossus activity for nasal sounds was found for some people, but not for others. Hence, the *basic* mechanism for lowering the velum might simply involve the *suppression* of activity of the levator palatini and rely on the restoring elastic force of the tissue to open the velopharyngeal port. This passive opening mechanism of the velopharyngeal port might simply be slower than the active muscle-controlled closing mechanism for some subjects, which could partly explain our results.

Stevens [24] suggested an *aerodynamic* reason that may explain higher velocities for velum raising than lowering (however without quantitative evidence): When a nasal is followed by an obstruent, the creation of the oral constriction or closure for the obstruent raises the intraoral pressure that exerts a force on the soft palate and so accelerates the raising movement. To find out whether or not this intraoral pressure increase really plays a role for faster raising movements, one could repeat the presented experiment with ingressive speech, where no intraoral pressure would build up. However, our preliminary attempts showed that it is very hard for most subjects to produce the timed phoneme changes with ingressive speech.

In summary, velocity differences between velum raising and lowering occur both in natural (fluent) speech, as indicated in previous studies, and in artificially timed speech, as in the present study. However, the effect size seems to be quite subject-dependent. While confounding contextual and prosodic factors could be minimized with the present experimental design, a discrimination of biomechanical and aerodynamic causes for the observed effect was not possible. In future work, it would be interesting to study potential differences between men and women with regard to velar movements, as our data suggest that the direction-dependent velocity differences were greater for the male speakers. Future work will also explore the potential benefit of including these velocity differences in articulatory speech synthesis. Finally, the correlation between LPW and velar movements should be further substantiated for a higher number of people to ensure that LPW velocity is a reliable surrogate of velar velocity.

Acknowledgments. This study was funded by the German Research Foundation (DFG), grant no. BI 1639/4-1.

References

1. Alfonso, P.J., Baer, T.: Dynamics of vowel articulation. Lang. Speech **25**(2), 151–173 (1982)
2. Amelot, A.: Etude aérodynamique, fibroscopique, acoustique et perceptive des voyelles nasales du français. Ph.D. thesis, Université de la Sorbonne nouvelle-Paris III (2004)

3. Amelot, A., Crevier-Buchman, L., Maeda, S.: Observations of the velopharyngeal closure mechanism in horizontal and lateral direction from fiberscopic data. In: 15th International Congress of Phonetic Sciences, Barcelona, Spain, pp. 3021–3024 (2003)
4. Bell-Berti, F.: Velopharyngeal function: a spatio-temporal model. In: Lass, N. (ed.) Speech and Language: Advances in Basic Research and Practice, pp. 291–316. Academic Press, New York (1980)
5. Benguerel, A.P., Hirose, H., Sawashima, M., Ushijima, T.: Velar coarticulation in French: a fiberscopic study. J. Phon. **5**(2), 149–158 (1977)
6. Birkholz, P.: Modeling consonant-vowel coarticulation for articulatory speech synthesis. PLoS ONE **8**(4), e60603 (2013)
7. Birkholz, P., Hoole, P.: Intrinsic velocity differences of lip and jaw movements: preliminary results. In: Proceedings of the Interspeech 2012, Portland, Oregon, USA, pp. 2017–2020 (2012)
8. Birkholz, P., Hoole, P., Kröger, B.J., Neuschaefer-Rube, C.: Tongue body loops in vowel sequences. In: 9th International Seminar on Speech Production (ISSP 2011), Montreal, Canada, pp. 203–210 (2011)
9. Dalston, R.M.: Photodetector assessment of velopharyngeal activity. Cleft Palate J. **19**(1), 1–8 (1982)
10. Fels, S., Lloyd, J.E., Van Den Doel, K., Vogt, F., Stavness, I., Vatikiotis-Bateson, E.: Developing physically-based, dynamic vocal tract models using ArtiSynth. In: Proceedings of the 7th International Seminar on Speech Production (ISSP 2006), Ubatuba, Brazil, pp. 419–426 (2006)
11. Fuchs, S., Perrier, P.: On the complex nature of speech kinematics. ZAS Pap. Linguist. **42**, 137–165 (2005)
12. Halle, M.: On distinctive features and their articulatory implementation. Nat. Lang. Linguist. Theory 91–105 (1983)
13. Horiguchi, S., Bell-Berti, F.: The velotrace: a device for monitoring velar position. Cleft Palate J. **24**(2), 104–111 (1987)
14. Iskarous, K., Nam, H., Whalen, D.H.: Perception of articulatory dynamics from acoustic signatures. J. Acoust. Soc. Am. **127**(6), 3717–3728 (2010)
15. Kleiner, C., Kainz, M.A., Echternach, M., Birkholz, P.: Speed differences in laryngeal adduction and abduction gestures. J. Acoust. Soc. Am. (submitted)
16. Krakow, R.A., Huffman, M.K.: Instruments and techniques for investigating nasalization and velopharyngeal function in the laboratory: an introduction. In: Nasals, Nasalization, and the Velum, pp. 3–59. Academic Press (1993)
17. Lukežič, A., Vojíř, T., Čehovin, L., Matas, J., Kristan, M.: Discriminative correlation filter tracker with channel and spatial reliability. Int. J. Comput. Vision **126**(7), 671–688 (2018)
18. Martins, P., Oliveira, C., Silva, S., Teixeira, A.: Velar movement in European Portuguese nasal vowels. In: Procedings of IberSPEECH 2012, Madrid, Spain, pp. 231–240 (2012)
19. Moll, K.L., Daniloff, R.G.: Investigation of the timing of velar movements during speech. J. Acoust. Soc. Am. **50**(2), 678–684 (1971)
20. Nam, H., Mooshammer, C., Iskarous, K., Whalen, D.: Hearing tongue loops: perceptual sensitivity to acoustic signatures of articulatory dynamics. J. Acoust. Soc. Am. **134**(5), 3808–3817 (2013)
21. Oliveira, C., Martins, P., Teixeira, A.: Speech rate effects on European Portuguese nasal vowels. In: Proceedings of the Interspeech 2009, Brighton, UK (2009)
22. Ryan, W.J., Hawkins, C.F.: Ultrasonic measurement of lateral pharyngeal wall movement at the velopharyngeal port. Cleft Palate J. **13**(2), 156–164 (1976)

23. Serrurier, A., Badin, P.: A three-dimensional articulatory model of the velum and nasopharyngeal wall based on MRI and CT data. J. Acoust. Soc. Am. **123**(4), 2335–2355 (2008)
24. Stevens, K.N.: Acoustic Phonetics. The MIT Press, Cambridge (1998)
25. Teixeira, A., Vaz, F.: European Portuguese nasal vowels: an EMMA study. In: Proceedings of the Eurospeech 2001, Aalborg, Denmark (2001)
26. Teixeira, A.J.S., Martinez, R., Silva, L.N., Jesus, L.M.T., Principe, J.C., Vaz, F.A.C.: Simulation of human speech production applied to the study and synthesis of European Portuguese. EURASIP J. Appl. Signal Process. **9**, 1435–1448 (2005)
27. Thiele, C., Mooshammer, C., Belz, M., Rasskazova, O., Birkholz, P.: An experimental study of tongue body loops in V1-V2-V1 sequences. J. Phon. **80**, 100965 (2020)
28. Xu, Y., Sun, X.: Maximum speed of pitch change and how it may relate to speech. J. Acoust. Soc. Am. **111**(3), 1399–1413 (2002)