IberSPEECH 2021
24-25 March 2021, Valladolid, Spain

The age effects on EP vowel production: an ultrasound pilot study

Luciana Albuquerque[1,2,3,4], *Ana Rita Valente*[1,3], *Fábio Barros*[1,3], *António Teixeira*[1,3], *Samuel Silva*[1,3], *Paula Martins*[1,5,6], *Catarina Oliveira*[1,5]

[1]Institute of Electronics and Informatics Engineering of Aveiro, Aveiro, Portugal
[2]Center for Health Technology and Services Research, University of Aveiro, Portugal
[3]Dep. Electronics, Telecommunications and Informatics, University of Aveiro, Portugal
[4]Department of Education and Psychology, University of Aveiro, Aveiro, Portugal
[5]School of Health Sciences, University of Aveiro, Portugal
[6]Institute of Biomedicine, University of Aveiro, Portugal
{lucianapereira, rita.valente, fabiodaniel, ajst, sss, pmartins, coliveira}@ua.pt

Abstract

For aging speech, there is a limited knowledge regarding the articulatory adjustments underlying the acoustic findings observed in previous studies. In this context, ultrasound imaging is a technology that can be safely used to study static and dynamic features of the articulators allowing comparisons of physiological differences between older and young adults during speech production. In order to investigate the age-related articulatory differences in European Portuguese vowels, the present study analyzes the tongue contours of the 9 European Portuguese oral vowels in isolated context and in a disyllabic sequence. From the tongue contours segmented from the Ultrasound images, several parameters were extracted (e.g., tongue height, tongue advancement) to allow comparisons between speakers of different age groups. For this study, while the analysis of data for more speakers is ongoing, we considered a set of four European Portuguese native female speakers of two different age groups and addressed the study of the oral vowels articulatory space. The preliminary results suggest that the vowel articulatory space, namely for disyllabic sequences, tend to be smaller in the older females.

Index Terms: 2D-Ultrasound tongue imaging, European Portuguese vowels, Speech Production, Aging speech

1. Introduction

The aging process causes specific alterations in the speech organs (e.g., stiffening of thorax, decreased lung capacity, weakening of respiratory muscles and atrophy of facial, mastication and pharyngeal muscles) [1, 2, 3], and all these changes are expected to play an important role in speech production. Over the years, although age-related variations on the acoustic properties of speech have been extensively investigated [2, 3, 4], its underlying articulatory details have not been well understood. On what concerns age-related changes on vowel formant frequencies (mostly F1 and F2), the results across studies are highly inconsistent [2, 3, 5, 6]. Previous acoustic results for European Portuguese (EP) were also not conclusive, as there are vowels that presented a different pattern of formant frequencies variation with age and gender [7, 8, 9]. [8, 10] observed that vowel formants tend to decrease mainly in females and to centralize in older males with aging, and these changes might be related to specific articulatory adjustments of the older speakers during speech.

Unlike acoustic studies, in which age-related speech variations have been widely studied since the 1960s [3], on what concerns articulatory studies there are only a few studies about effects of aging on tongue position and strength during speech production [11, 12, 13, 14]. Ultrasound (US) tongue imaging synchronized with audio can be used to investigate the physiological differences between old and young adult speech. For this reason, the main objective of the ongoing study is to investigate the age-related articulatory differences in EP vowels with US imaging. Additionally, since there is a paucity of literature on EP oral vowel production, the available data were collected mainly for acoustic studies [7, 8, 15, 16] or in articulatory studies of nasal vowels [17, 18, 19], this study also provides valuable insights to a first articulatory description of all EP oral vowels with US.

Due to the limited knowledge of the articulatory basis of previously acoustic findings concerning aging speech, the aim of this pilot study is to analyze and compare the US tongue contours of the 9 EP oral vowels in isolated context and in a disyllabic sequence produced by four female speakers between 19 and 66 years, to infer preliminary results of the age effect on EP vowel production, which is essential for the development of automatic speech recognition (ASR) systems suitable for older speech, and for clinical assessment and treatment of speech disorders.

2. Background and Related Work

Speech production has been studied essentially through speech sound acoustics (for EP: [7, 8, 15, 16]). Although frequencies are an indirect measure of articulatory movements, direct measures can be obtained over different articulatory techniques, such as electromagnetic articulography (EMA), real-time magnetic resonance (RTMRI) and US imaging (for EP: [17, 18, 19]). US imaging presents several advantages in comparison with the other referred techniques: it is a non-invasive, safe, portable and fast technology that is commonly used to demonstrate the midsagittal surface contour of the tongue [20], and it can contribute with important information for different areas in speech research [21, 22].

Despite US being a more affordable alternative for several contexts, enabling the acquisition of larger datasets, it demands adequate computational approaches for processing and analysis. US artifacts, corrupting noise, and the presence of spurious edges are also challenges for the processing of US images [22, 23]. Furthermore, a challenge in measuring US images is the lack of a physiological reference [23]. Even though the tongue contours are visible, there are no hard structure references (i.e., US does not image internal articulators other than the tongue), making it difficult to determine an exact position for the tongue in the vocal tract [24]. The head and transducer holders help overcome these problems [23], but cannot be guaranteed to be re-fitted to

10.21437/IberSPEECH.2021-52

the same location on a speakers' head in different trials [25]. For that, re-orienting images to a common co-ordinate system with a bite plane allows for some degree of normalization and it is more tolerant to error in the placing of the probe outwith the midsagittal plane [25].

The lack of reference points and the anatomical differences also introduce a difficulty in comparing speech data across speakers, that is, comparing lingual articulation across age groups raises problems of normalization [24, 26]. This issue has been also recognized in the acoustic studies of vowel formants; and indeed, the length of the vocal tract can be considered a greater influence in the articulatory field, leading to the need of normalization procedures [26]. External references or the image of internal articulators have been proposed as solutions [24, 23]. However, there is no commonly accepted method for comparing tongue shapes among speakers [27].

Several research studies developed since 2010 have been used US tongue images to investigate, on different language systems, the articulatory correlates of vowels production [28, 29]. Concerning the articulatory parameters, tongue contour and also height and advancement of the highest point of the tongue had been successfully used in previous US studies with vowels.

To the best of our knowledge, the articulatory studies concerning vowels' properties across lifespan are scarce and the majority focuses on coarticulation issues. An articulatory study with 3D electromagnetic articulography suggested that especially the tongue body was affected by age, and the movements for the vowels were slower in the older speakers compared to the younger ones [11]. The results of an US study of antecipatory velar-vowel coarticulation and speech stability in speakers who do and do not stutter across lifetime [13] indicated an age effect, with progressive less coarticulation and increase speech stability with aging. Regarding the performance on motor tasks with increasing age, a generalized slowness, decrease coordination, a diminished performance level and on precise motor control were observed [13, 11], which could affect vowel production.

3. Method

This cross-sectional study was approved by the Ethics Committee of Escola Superior de Enfermagem de Coimbra, Portugal (approval number 639/12-2019), and all participants agreed and signed the consent form before participating in the study.

3.1. Speakers and Corpus

Ultrasound data were collected from a convenience sample of four EP native female speakers, two young (S1 - 19 years; S2 - 31 years) and two old females (S3 - 63 years; S4 - 66 years), to test ages with more expected distinctive characteristics. Data collection took place in a pandemic year, which conditioned the acquisition of a larger number of participants, namely older people. All of them were in good health and with no reported history of neurological disorders or diseases, or any speech, language or hearing difficulties. Considering the speakers' anatomical characteristics, S1 speaker is 174 cm tall and weighs 95 kg; S2 is 171 cm tall and weighs 56 kg; S3 is 152 cm tall and weighs 59 kg; and S4 is 150 cm tall and weighs 52 kg.

The corpus consisted of all EP oral vowels [i], [e], [ɛ], [a], [o], [ɔ], [u], [ɨ] and [ɐ] in pseudoword context and in isolated context. The pseudoword list contained ˈpV.Cv sequences (started with the labial voiceless stop consonant [p]), where C was balanced for the place of articulation using the voiceless stop consonants [p], [t] and [k], and V was all EP oral vowels in stressed position. The last vowel (i.e., v) corresponds only to the vowels [u], [ɨ] and [ɐ]. The stimuli were embedded in a carrier sentence "Em pVCv temos V" (*In pVCv we have V.*). For each vowel, three different pseudowords were selected. Each carrier sentence was repeated 3 times. Thus, each speaker produced 81 individual utterances, i.e., 9 repetitions of each vowel, per context.

3.2. Data Acquisition

The participants were asked to seat, facing a computer screen displaying prompts, and to wear a stabilization helmet [30], in order to ensure that neither the speaker's head nor the transducer moved during the experiment.

Synchronous acquisition of US images and speech sounds through Articulate Assistant Advanced software (AAA) [31] took place in quiet rooms using an endocavitary probe (65EC10EA) with 90º field of view positioned under the participants' chin. US images were collected with an US machine Mindray DP6900 at a frame rate of 60 Hz and the depth setting was 97 mm. Audio was collected with a Philips SBC ME400 microphone connected to an external sound system (UA-25 EX USB). The recorded data was collected as video and audio synchronized with a SyncBrightUp unit [32].

Instructions were provided prior to recording to ensure familiarity with the speech materials. The speech material was presented in three randomized blocks (i.e., front ([i], [e], [ɛ]), central ([ɨ], [ɐ], [a]) and back ([u], [o], [ɔ]) vowels). Each block began and finished with the production of the sequence /tatatata/ to assess sound and image synchronization. Also, at the start of each block a recording of the bite plane was obtained in order to image the occlusal plane, which is a reliable method for the definition of horizontal and vertical orientations in the vocal tract [25, 33]. That is, the speaker was asked to bite and press their tongue against a flat plastic plate, which results in their tongue bulging upward at the back edge of the bite plate [25, 33].

Taking into consideration different anatomical characteristics of each speaker and that the optimal probe orientation is vowel dependent, its orientation was adjusted, along the sessions, for each block of vowels, to enable the best possible imaging of the tongue. The bite plane sequences were then used as a common referential, for each speaker.

3.3. Data processing and statistical analyses

Data processing – The acoustic files were exported from AAA software in WAV format for automatic segmentation at word and phoneme level using WebMAUS [34], and then imported into Praat software [35], to manually check the accuracy of the vowel boundaries.

To reduce the impact of the noisy nature of the US images, a pre-processing was applied and consists in: 1) cropping the US image to remove irrelevant information and to select a region of interest (ROI); 2) applying a phase symmetry filter to the ROI to enhance the outline corresponding of the tongue surface [36, 37]; 3) applying a radial sweep approach with 5º of angular distance [29]; 4) collecting of all pixel intensities; 5) extraction of the highest intensity point for each radial sweep. More details in [38, 39]. Given the challenging nature of the images, and to ensure the reliability of the data, the tongue segmentation of all frames of the vowel occurrences was revised by three annotators with experience in speech production analysis (see Fig. 1).

For each block of vowels, tongue contours were rotated to the speaker's bite plane obtained in the corresponding block, so that the image of the occlusal plane was observed to be parallel to the upper and lower edges of the video pane [25, 26, 33].

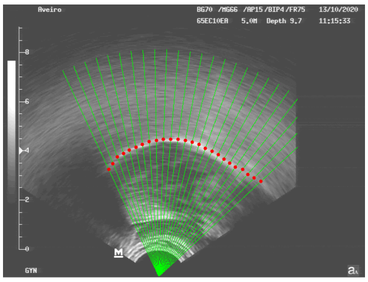

Figure 1: *Illustrative US image with radial sweeps and the extracted points on the tongue contour for vowel [a]*

Data were exported after rotation and the origin of the coordinates system corresponds to the back of the bite plate, 4 cm from the upper incisors along the occlusal plane (see the bite plane traces in Fig. 2).

For the present study, only the tongue contour of the temporal midpoint of the vowels, which consistently contained an articulatorily steady part of the vowel [33], were exported in cartesian coordinates. Ideally, the total number of vowels analyzed per speaker was 81, which corresponds to 81 frames. However some frames had to be discarded from the analysis due to poor US image quality for some speakers (namely for S3), vowels (mostly back vowels), and/or context (mostly in isolated context). Furthermore, due to the fact that some tongue contours were not totally visible, some tokens were segmented incomplete and not being able to find important descriptors for vowel analysis, such as the highest point of the tongue body.

Articulatory Measures – Each vowel tongue contours per speaker and context were obtained as the median value of all vowel occurrences. The highest point of the tongue's contour was extracted and represents the highest point of the tongue body (i.e., tongue height, TH). The x-coordinate reflects the front back position of the tongue in the y coordinate (i.e., tongue advancement, TA). Pixel to cm conversion was made considering 1 cm corresponds to 44 pixels in the image. The TH and TA for each vowel, per context and speaker, was obtained based on the median value of all tokens at the temporal midpoint, which reduces the effect of the flanking consonants and the effect of measurement errors.

4. Results

In this section a summary of the main findings for the analysis of the vowel tongue configuration on EP vowels are presented. Also some considerations about inter and intra-speaker differences are reported. Due to space limitations, contours and vowel articulatory spaces of all speakers are not given here, and only relevant differences are presented. Note that, for speaker S3, none completed contour of the back vowels [o] and [ɔ] in isolated context was obtained, which affect the TH and TA measures presented for this speaker.

Tongue contours – Concerning the tongue contours analysis, results for one speaker, both contexts, are presented in Fig. 2. The differences in tongue contours appear to be clearly vowel dependent in isolated vowels. For vowels in pVCv sequences the contours for different vowels tend to be more similar.

Articulatory Measures – Considering the tongue contours, the highest point of the tongue body was determined and analyzed. Fig. 3 summarizes the TH and TA values obtained for each speaker by vowel in pVCv sequence, and in isolated context. In general, vowels in pVCv sequence presented higher variability of TA and TH values than isolated vowels due to the influence of the consonantal context [40, 41]. Regardless of the vowel

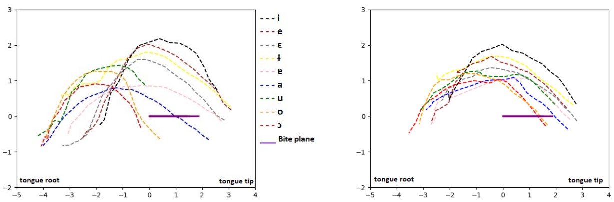

Figure 2: *Tongue contours for all vowels in isolated (left side) and in pVCv sequences (right side) for speaker S4 (scale in cm)*

context, for each speaker, the higher TH was observed for vowel [i], and the lowest TH was obtained, mostly, for the vowel [a]. The lowest TA was obtained, mostly, for the vowel [o]. Concerning the vowel that presented the highest TA, data showed more variability, with some speakers present vowel [ɐ] with higher TA than expected, mostly in isolated context. Concerning the highest TA, as data showed greater variability, no vowel can be pointed.

Articulatory vowel space – For intra-speaker comparisons, Fig 4 represents the articulatory space defined by TA and TH of the cardinal EP oral vowels ([a], [i] and [u]) for each speaker, in isolated and in pVCv context. Each vowel is represented by the median TH and TA. Comparing the TH and TA of the tongue contours of these female speakers, it can be observed that articulatory space tends to be smaller when vowels occur in pVCv sequences comparing with isolated vowels. The data also tends to indicate that the vowel articulatory space differences observed between both contexts is higher for the old females. That is, the articulatory vowel space area, namely for pVCv sequences, tend to be smaller in the older females.

Articulation variability – Plots in Fig. 4 only provide the average, but variability of productions is also very important to analyze. Fig. 5 represents individual productions and information regarding dispersion based on ellipses. Fig. 5 represents the TH and TA of the total number of occurrences of the cardinal vowels in both contexts for two speakers (one young and one old). In those graphs it can observed that the dispersion of isolated vowels is lower than in pVCv sequences, mainly for vowel [a]. Also, in both contexts, for these two speakers in can be observed that vowel [a] presents the highest variability.

5. Conclusions

This paper presents the initial results of the automatic extraction contours and the determination of the highest position of the tongue body for young and old Portuguese females. The method revealed being able to detect articulatory measures for young and old speakers, making possible the production of the first vowel articulatory space representation for EP. Regarding limitations, this study presents a reduced sample that, due non-normalization procedures, hinder inter-speaker comparisons. Also, the noisy nature of the images make the segmentation demanding and could difficult the accurate determination of articulatory measures for some vowels. These issue could be reflected in the greater variability observed for certain vowels.

The vowel articulatory space reduction observed for isolated vowels in comparison with the vowels in pVCv sequences might be related with the tendency to hyperarticulate isolated vowels. So, this vowel articulatory space reduction was similar between vowels in clear speech versus in conversational speech [40], or in long vowels versus short vowels [42], for other languages. Also the tendency to hyperarticulate isolated vowels might be in the origin of the more distinct tongue contours for these vowels.

Future work – Ensure inter-speaker comparison through the aplication of normalization procedures; improvement of

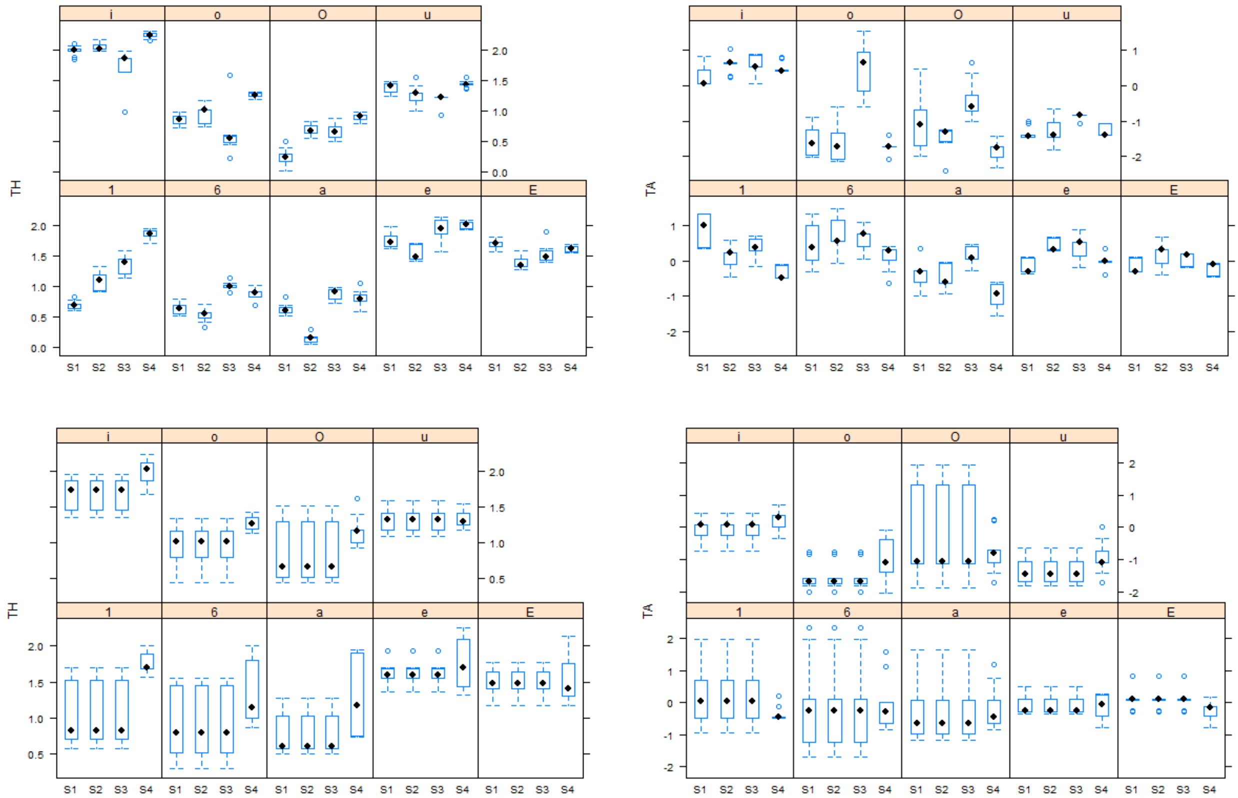

Figure 3: *TH e TA by vowel type and speaker (SAMPA notation). Top: isolated vowels; Bottom: vowels in pVCv sequence (scale in cm)*

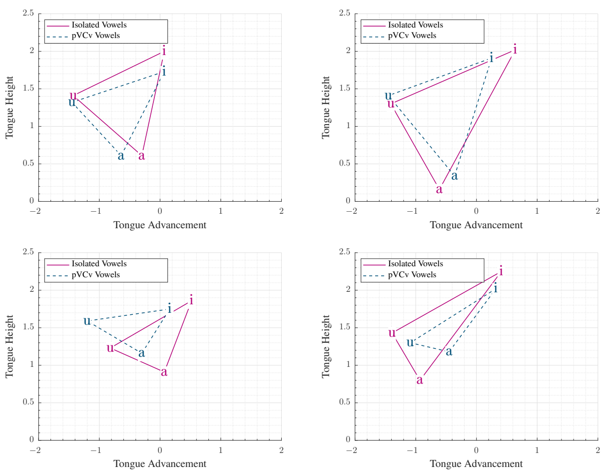

Figure 4: *Articulatory vowel space of the EP cardinal vowels in isolated (pink solid lines) and in pVCv context (blue dashed lines). Top: young females (S1 and S2); bottom: old females (S3 and S4)*

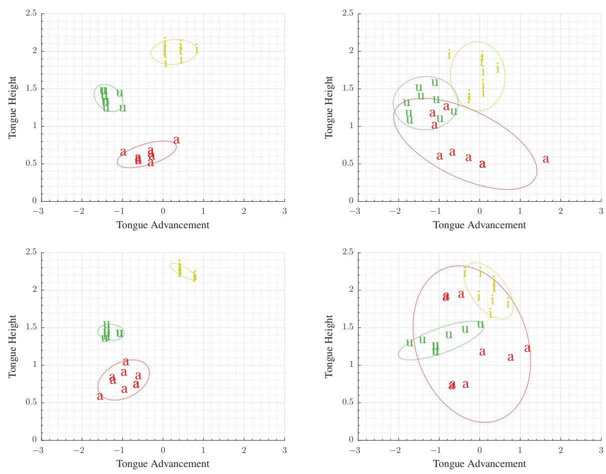

Figure 5: *Vowel articulatory cluster size of the EP cardinal vowels of two speakers in isolated vowels (left side) and in pVCv sequences (right side). Top: young female (S1); Bottom: old female (S4)*

the determination of the highest point of the tongue; explore SSANOVA to study vowel contours; investigate articulatory movement and velocity; a large study of the age effect on vowel articulation measures.

6. Acknowledgements

This research was financially supported by the projects VoxSenes (POCI-01-0145-FEDER-03082) and MEMNON (POCI-01-0145-FEDER-028976) – COMPETE2020 under POCI and FEDER, and by national funds (OE), through FCT/MCTES –, SOCA – Smart Open Campus CENTRO-01-0145-FEDER-000010 (Portugal 2020 under POCI and FEDER) and by IEETA Research Unit funding (UIDB/00127/2020). First author was funded by FCT grant SFRH/BD/115381/2016.

7. References

[1] P. Massimo and P. Elisa, “Age and Rhtymic Variations: A study on Italian,” in *INTERSPEECH*, Singapore, 2014, pp. 1234–1237.

[2] S. E. Linville, *Vocal aging*. Australia, San Diego: Singular Thomson Learning, 2001.

[3] S. Schötz, *Perception, analysis and synthesis of speaker age*. Lund University: Linguistics and Phonetics, 2006, vol. 47.

[4] R. Vipperla, S. Renals, and J. Frankel, "Ageing voices: The effect of changes in voice parameters on ASR performance," *EURASIP J. Aud. Speech Music Process*, pp. 1–10, 2010.

[5] M. P. Rastatter, R. A. McGuire, J. Kalinowski, and A. Stuart, "Formant frequency characteristics of elderly speakers in contextual speech," *Folia Phoniatrica et Logopaedica*, vol. 49, no. 1, pp. 1–8, 1997.

[6] J. T. Eichhorn, R. D. Kent, D. Austin, and H. K. Vorperian, "Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women," *Journal of Voice*, vol. 32, no. 5, pp. 644.e1–644.e9, 2018.

[7] L. Albuquerque, C. Oliveira, A. Teixeira, P. Sa-Couto, J. Freitas, and M. S. M. Dias, "Impact of age in the production of European Portuguese vowels," in *INTERSPEECH*, Singapore, 2014, pp. 940–944.

[8] L. Albuquerque, C. Oliveira, A. Teixeira, P. Sa-Couto, and D. Figueiredo, "A comprehensive analysis of age and gender effects in European Portuguese oral vowels," *Journal of Voice*, no. In press, dec 2020.

[9] T. Pellegrini, A. Hämäläinen, P. B. de Mareüil, M. Tjalve, I. Trancoso, S. Candeias, M. S. Dias, and D. Braga, "A corpus-based study of elderly and young speakers of European Portuguese: acoustic correlates and their impact on speech recognition performance," in *INTERSPEECH*, Lyon, 2013, pp. 852–856.

[10] L. Albuquerque, C. Oliveira, A. Teixeira, P. Sa-Couto, and D. Figueiredo, "Age-related changes in European Portuguese vowel acoustics," in *INTERSPEECH*, Graz, 2019, pp. 3965–3969.

[11] A. Hermes, J. Mertens, and D. Mücke, "Age-related Effects on Sensorimotor Control of Speech Production," in *INTERSPEECH*, Hyderabad, 2018, pp. 1526–1530.

[12] P. De Decker and S. Mackenzie, "Tracking the phonological status of /l/ in Newfoundland English: Experiments in articulation and acoustics," *J. Acoust. Soc. Am.*, vol. 142, no. 1, pp. 350–362, 2017.

[13] A. J. Belmont, "Anticipatory Coarticulation and Stability of Speech in Typically Fluent Speakers and People Who Stutter Across the Lifespan: An Ultrasound Study," Master of Science, University of South Florida, 2015.

[14] A. T. Neel and P. M. Palmer, "Is Tongue Strength an Important Influence on Rate of Articulation in Diadochokinetic and Reading Tasks?" *JSLHR*, vol. 55, pp. 235–246, 2012.

[15] P. Escudero, P. Boersma, A. S. Rauber, and R. A. H. Bion, "A cross-dialect acoustic description of vowels: Brazilian and European Portuguese," *J. Acoust. Soc. Am.*, vol. 126, no. 3, pp. 1379–1393, 2009.

[16] C. Oliveira, M. M. Cunha, S. Silva, A. Teixeira, and P. Sa-Couto, "Acoustic analysis of European Portuguese oral vowels produced by children," in *IberSPEECH*, vol. 328, Madrid, 2012, pp. 129–138.

[17] C. Oliveira, P. Martins, S. Silva, and A. Teixeira, "An MRI study of the oral articulation of European Portuguese nasal vowels," in *13th INTERSPEECH*, Portland, OR, USA, 2012, pp. 2690–2693.

[18] C. Cunha, S. Silva, A. Teixeira, C. Oliveira, P. Martins, A. Joseph, and J. Frahm, "On the Role of Oral Configurations in European Portuguese Nasal Vowels," in *INTERSPEECH*, Graz, Austria, 2019, pp. 3332–3336.

[19] C. Oliveira, P. Martins, and A. Teixeira, "Speech rate effects on European Portuguese nasal vowels." in *INTERSPEECH*, Brighton, 2009, pp. 480–483.

[20] L. Lancia, P. Rausch, and J. S. Morris, "Automatic quantitative analysis of ultrasound tongue contours via wavelet-based functional mixed models," *J. Acoust. Soc. Am.*, vol. 137, no. 2, pp. EL178–EL183, 2015.

[21] M. H. Mozaffari, S. Wen, N. Wang, and W. Lee, "Real-time automatic tongue contour tracking in ultrasound video for guided pronunciation training," in *14th VISIGRAPP*, vol. 1, 2019, pp. 302–309.

[22] Y. S. Akgul, C. Stone, and K. Maureen, "Automatic extraction and tracking of contours," *TRANSACTIONS ON MEDICAL IMAGING*, vol. 18, no. 10, pp. 1035–1045, 1999.

[23] M. Stone, "A guide to analysing tongue motion from ultrasound images," *Clin Linguist Phon*, vol. 19, no. 6-7, pp. 455–501, 2005.

[24] N. Zharkova, N. Hewlett, and W. J. Hardcastle, "Coarticulation as an indicator of speech motor control development in children: An ultrasound study," *Motor Control*, vol. 15, no. 1, pp. 118–140, 2011.

[25] J. M. Scobbie, E. Lawson, S. Cowen, J. Cleland, and A. A. Wrench, "A common co-ordinate system for mid-sagittal articulatory measurement," in *QMU CASL Working Papers WP-20*, Edinburgh, 2011.

[26] P. Strycharczuk and J. M. Scobbie, "Fronting of Southern British English high-back vowels in articulation and acoustics," *J. Acoust. Soc. Am.*, vol. 142, no. 1, pp. 322–331, 2017.

[27] G. Barbier, P. Perrier, L. Ménard, Y. Payan, M. Tiede, and J. Perkell, "Speech planning in 4-year-old children versus adults: Acoustic and articulatory analyses," in *INTERSPEECH*, 2015.

[28] K. Comivi Alowonou, J. Wei, W. Lu, Z. Liu, K. Honda, and J. Dang, "Acoustic and Articulatory Study of Ewe Vowels: A Comparative Study of Male and Female," in *INTERSPEECH*. Graz, Austria: ISCA, 2019, pp. 1776–1780.

[29] L. Ménard, C. Toupin, S. R. Baum, S. Drouin, J. Aubin, and M. Tiede, "Acoustic and articulatory analysis of French vowels produced by congenitally blind adults and sighted adults," *J. Acoust. Soc. Am.*, vol. 134, no. 4, pp. 2975–2987, 2013.

[30] Articulate Instruments Ltd., "Ultrasound stabilisation headset users manual," Edinburgh, UK, 2008.

[31] Articulate Assistant Ltd., "Articulate Assistant Advanced ultrasound module user manual," 2014.

[32] Articulate Instruments Ltd., "SyncBrightUp users manual," Edinburgh, UK, 2010.

[33] M. Dokovova, M. Sabev, J. M. Scobbie, R. Lickley, and S. Cowen, "Bulgarian vowel reduction in unstressed position: an ultrasound and acoustic investigation," in *19th ICPhS*, 2019, pp. 2720–2724.

[34] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech and Language*, vol. 45, pp. 326–347, 2017.

[35] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," University of Amsterdam, 2012.

[36] P. Kovesi *et al.*, "Symmetry and asymmetry from local phase," in *Tenth Australian joint conference on artificial intelligence*, vol. 190. Citeseer, 1997, pp. 2–4.

[37] E. Karimi, L. Ménard, and C. Laporte, "Fully-automated tongue detection in ultrasound images," *Computers in Biology and Medicine*, vol. 111, no. 103335, pp. 1–13, 2019.

[38] F. Barros, A. R. Valente, L. Albuquerque, S. Silva, A. Teixeira, and C. Oliveira, "Contributions to a quantitative unsupervised processing and analysis of tongue in ultrasound images," *Lecture Notes in Computer Science*, vol. 12132 LNCS, pp. 170–181, 2020.

[39] F. Barros, S. Silva, L. Albuquerque, A. R. Valente, A. Teixeira, P. Martins, and C. Oliveira, "Towards the use of ultrasonography to study aging effects in vowel production," in *12th ISSP*, 2020.

[40] J. Y. Song, "The use of ultrasound in the study of articulatory properties of vowels in clear speech," *Clin Linguist Phon*, vol. 31, no. 5, pp. 351–374, 2017.

[41] N. Zharkova, F. E. Gibbon, and W. J. Hardcastle, "Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation," *Clinical linguistics & phonetics*, vol. 29, no. 4, pp. 249–265, 2015.

[42] W.-S. Lee, "Articulatory–Acoustical Relationship in Cantonese Vowels," *Language and Linguistics*, vol. 17, no. 4, pp. 477–500, 2016.