

## Article

# Exploring the Age Effects on European Portuguese Vowel Production: An Ultrasound Study

Luciana Albuquerque <sup>1,\*</sup> , Ana Rita Valente <sup>2</sup> , Fábio Barros <sup>2</sup> , António Teixeira <sup>2</sup> , Samuel Silva <sup>2</sup> ,  
Paula Martins <sup>3</sup>  and Catarina Oliveira <sup>4</sup> 

<sup>1</sup> Institute of Electronics and Informatics Engineering of Aveiro (IEETA), CINTESIS.UA, Department of Education and Psychology, University of Aveiro, 3810-193 Aveiro, Portugal

<sup>2</sup> Institute of Electronics and Informatics Engineering of Aveiro (IEETA), Department Electronics, Telecommunications and Informatics (DETI), University of Aveiro, 3810-193 Aveiro, Portugal; rita.valente@ua.pt (A.R.V.); fabiodaniel@ua.pt (F.B.); ajst@ua.pt (A.T.); sss@ua.pt (S.S.)

<sup>3</sup> Institute of Electronics and Informatics Engineering of Aveiro (IEETA), Institute of Biomedicine, School of Health Sciences (ESSUA), University of Aveiro, 3810-193 Aveiro, Portugal; pmartins@ua.pt

<sup>4</sup> Institute of Electronics and Informatics Engineering of Aveiro (IEETA), School of Health Sciences (ESSUA), University of Aveiro, 3810-193 Aveiro, Portugal; coliveira@ua.pt

\* Correspondence: lucianapereira@ua.pt

**Abstract:** For aging speech, there is limited knowledge regarding the articulatory adjustments underlying the acoustic findings observed in previous studies. In order to investigate the age-related articulatory differences in European Portuguese (EP) vowels, the present study analyzes the tongue configuration of the nine EP oral vowels (isolated context and pseudoword context) produced by 10 female speakers of two different age groups (young and old). From the tongue contours automatically segmented from the US images and manually revised, the parameters (tongue height and tongue advancement) were extracted. The results suggest that the tongue tends to be higher and more advanced for the older females compared to the younger ones for almost all vowels. Thus, the vowel articulatory space tends to be higher, advanced, and bigger with age. For older females, unlike younger females that presented a sharp reduction in the articulatory vowel space in disyllabic sequences, the vowel space tends to be more advanced for isolated vowels compared with vowels produced in disyllabic sequences. This study extends our pilot research by reporting articulatory data from more speakers based on an improved automatic method of tongue contours tracing, and it performs an inter-speaker comparison through the application of a novel normalization procedure.

**Keywords:** 2D-ultrasound tongue imaging; European Portuguese vowels; speech production; aging speech



**Citation:** Albuquerque, L.; Valente, A.R.; Barros, F.; Teixeira, A.; Silva, S.; Martins, P.; Oliveira, C. Exploring the Age Effects on European Portuguese Vowel Production: An Ultrasound Study. *Appl. Sci.* **2022**, *12*, 1396. <https://doi.org/10.3390/app12031396>

Academic Editor: Valentino Santucci

Received: 30 December 2021

Accepted: 25 January 2022

Published: 28 January 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Aging involves changes at physiological, cognitive, psychological, and social levels. Age impacts the human body in a plethora of ways, entailing changes to the musculo-skeletal, the cardiovascular, the respiratory, and the central nervous systems [1], and the speech production system is no exception [2]. Thus, the aging process causes specific alterations in the speech organs (e.g., decreased lung capacity, weakening of respiratory muscles and atrophy of facial, mastication, and pharyngeal muscles) [3–5], and all these changes are expected to play an important role in speech production. Although limited research has been carried out specifically to ascertain what changes occur in the supralaryngeal structures (particularly in the tongue) with aging, modifications in these structures could considerably affect the speech production [6].

Although age-related variations on the acoustic properties of speech have been extensively investigated over the years [4,5,7], its underlying articulatory adjustments have not been well understood. Due to the fact that the vowel formant values are suited to

articulatory interpretations (i.e., formant frequencies reflect the length and configuration of the vocal tract [8–12]), formant measurements have been used in the study of speech production. Thus, the first (F1) and second (F2) formant frequencies reflect primarily tongue position and lips rounding [13]. As for age-related changes on vowel formant frequencies (mostly F1 and F2), the results across studies are highly inconsistent [4,5,11,14]. In addition, European Portuguese (EP) vowels presented a different pattern of formant frequencies variation with age and gender [15–17]. Albuquerque et al. [16,18] observed that vowel formants tend to decrease mainly in females and to centralize in males with aging, and these changes might be related to specific articulatory adjustments of the older speakers during speech.

Unlike acoustic studies, for articulatory studies, there are only a few that examine the effects of aging on tongue position and strength during speech production [1,19–22]. Ultrasound (US) tongue imaging synchronized with audio can be used to investigate the physiological differences between old and young adult speech [19,23], as approached in our previous work with four female speakers [22]. Nonetheless, US tongue image processing and the inter-speaker comparison are challenging [22]. Targeting the main limitations of our previous study, the main purpose of this work is to perform a larger study of the age-related articulatory differences in EP vowels with US imaging and to investigate normalization procedures. Additionally, since there is a paucity of literature on EP oral vowels production, and the available data were collected mainly for acoustic studies [15,16,24,25] or for articulatory studies of nasal vowels [26–28], this study also provides valuable insights for the accurate articulatory description of EP oral vowels with US.

The data obtained will allow inferring 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.

### 1.1. Background

#### 1.1.1. Aging Effects on the Articulatory Subsystem

Noticeable anatomic and physiologic changes in the vocal tract or supralaryngeal system have been reported from young adulthood to old age [29]. During this period, facial bones continue to grow (3–5%) [29]. Additionally, a slight lowering of the larynx was observed, which is due to the atrophy of the head and neck muscles as well as the thinning of the intervertebral disks (more common in females), which leads to an increase of the length of the vocal tract [4,5,30,31].

In general, changes observed in the vocal tract include atrophy of the facial, the mastication, and the pharyngeal muscles, a decrease of the soft-palate volume, and loss of lip strength [3–5,7,29,32–34]. The loss of elasticity and the thinning of the oral mucosa with the deterioration of attachments of the epithelium and connective tissue to bone are more apparent after age 70; however, these changes may happen earlier in life [29,33]. Extensive degenerative changes occur in the temporomandibular joint (gradual reduction in size and reductions in blood supply), which controls the jaw movement during speech production [5,7,29]. In addition, a diminished accuracy of the lower lip and jaw when performing rapid movements has been reported with age [3,32]. Furthermore, the shape of the oral cavity may change with the loss of teeth and the introduction of dentures [35].

The salivary glands undergo acinar atrophy, ductal irregularities, and parenchymal replacement by fibrous and/or adipose tissue with aging, which causes a decline in salivary function [33,34].

Regarding the tongue, age-related muscular changes include muscle atrophy [36], decreased thickness of the lingual epithelium, reduced muscle fiber diameter, and fissuring of the tongue surface [29,37]. Concerning the size of the tongue, contradictory findings have been reported in the literature [6]. While Sonies et al. [23] observed that the tongue of older people was reduced in thickness during rest, other studies have suggested that

age-related changes in the tongue size are minimal to non-existent [6,34,37], which might be explained by the large increase of deposits of fatty tissue in the tongue to compensate the loss of muscle fibers [6,36,37]. On the other hand, Klein [33] suggested an increase in tongue size because of two factors: muscle tone loss and the need to expand to fill the oral cavity space, as teeth are lost with aging.

From a functional perspective, the older individuals have been reported to experience significant declines in tongue strength, although endurance remains stable throughout the major part of life [29,38,39]. Mortimore et al. [40,41] observed that the maximal tongue protrusion force decreases with age, while the tongue fatigability increases. In addition, drying of the tongue and of the oral cavity may lead to an increased resistance of tongue movement across the palate during speech movements [6].

Furthermore, due to the loss of flexibility and muscular strength, aging can lead to deficits in speech motor control [1,20,42]. Previous studies about age effects in articulatory motor performance have demonstrated slower movements and reduced regularity in the rhythm of tongue movements, reduced tongue retraction, and smaller reductions in distance traveled by the tongue during fast speech rates with age [1,6,23,37]. Chantaramanee et al. [43] showed that the quality of the tongue muscles in healthy older speakers, assessed by the echo intensity of the tongue, is an indicator of tongue thickness at the middle and base of the tongue.

All of these changes affect the vowel production [6,7], as changes in vocal tract dimensions and in tongue motion may influence the resonance patterns in older speakers [7,44].

#### 1.1.2. Articulatory Studies of Age Effects

Direct measures of articulatory movements can be obtained over different techniques, such as electropalatography (EPG), electromagnetic articulography (EMA), real-time magnetic resonance (RTMRI), and US imaging.

To the best of our knowledge, the articulatory studies concerning vowels' properties across lifespan are scarce, and the majority of them focus on coarticulatory issues [20,45]. An articulatory study with 3D EMA for German suggested that the tongue body was especially affected by age, and the movements for the vowels were slower in the older speakers compared to the younger ones [1]. The results of a US study of anticipatory velar–vowel coarticulation and speech stability in English speakers who stutter and do not stutter across lifetime [20] indicated an age effect, with progressive less coarticulation and an increase of speech stability with aging. With US, Sonies et al. [23] studied the tongue movements in healthy older and younger English speakers, during the repetition of the phonemes [a], [i], and [k], and they observed a reduction in tongue retraction during the vowel [a] production for older speakers. A pilot study for EP with US [22] suggested that the vowel articulatory space tends to be smaller in the older EP females comparing with younger females.

Regarding consonants, in a US study for Newfoundland English, De Decker and Mackenzie [19] found a significant effect of age on the articulatory properties of /l/ in Newfoundland English, with older speakers being more likely to exhibit distinctions in tongue gestures between the word-initial and word-final positions.

#### 1.1.3. Ultrasound Imaging of Speech

US imaging presents several advantages in comparison with other articulatory techniques (e.g., EPG, EMA, RTMRI): it is a non-invasive, safe, portable, and fast technology that is commonly used to image the midsagittal surface contour of the tongue [46,47], and it can contribute with important information in different areas in speech research [48,49].

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, the presence of spurious edges, and the lack of a physiological reference are also challenges for the processing of US images [49,50]. 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 [47,51]. The head and transducer holders help overcome these problems [50], but they cannot be guaranteed to be re-fitted to the same location on a speaker's head in different trials [52]. For that, re-orienting images to a common coordinate 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 [52].

Furthermore, often, the US tongue contours do not provide information on the tongue tip and/or on the tongue root due to the interfering presence of the mandible and hyoid bones [47]. Thus, the US image segmentation is strongly influenced by the quality of data [53]. The speaker size and age also tend to influence the quality of the US tongue image (i.e., women and children tend to produce better images than men, younger subjects generally image better than older subjects, and thinner subjects tend to produce better images than larger subjects) [47]. Tabain [47] suggests that these trends may be related to fat levels in the tongue and moisture levels in the mouth.

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 [51,54]. However, there is no commonly accepted method for comparing tongue shapes among speakers [55].

Several research studies have been using US tongue images to investigate, on different languages, the articulatory correlates of vowel production [54,56,57]. Most of them have focused on coarticulation aspects among children, adolescents, and adults [20,55,58,59]. Concerning the articulatory parameters, tongue contour, height, and advancement of the highest point of the tongue (i.e., TH and TA), lengths of posterior tongue surface (LPTS) and anterior oral cavity (LAOC) have been successfully used in previous US studies with vowels [60–63]. In addition, ratios such as Dorsum Excursion Index (DEI—quantifies the extent of tongue dorsum excursion toward the palate during an articulation) and Tongue Constraint Position Index (TCPI—represents the place of maximal constriction) have been used in coarticulation studies [59,64], which allow the comparison across speakers without the need to perform a normalization for vocal tract size [65]. Nonetheless, to obtain these indices, it is essential to ensure that the tongue curve in the image is visible between the mandible and hyoid bones shadows, and that both shadows are also visible [65] (as well as to obtain the LPTS and LAOC measures [60]).

## 2. Materials and Methods

### 2.1. Participants

Speech production data were collected from a convenience sample of 10 EP native female speakers, 5 young (between 23 and 32, with a mean age of 27.0) and 5 old (between 59 and 73, with a mean age of 62.4) females. These participants were a subset of 32 speakers recorded for a larger study. They were selected due to the high quality of their US data (which was crucial to the articulatory analysis [54]), and only one gender was selected to avoid possible confounding effects of gender [6].

Participants were recruited through personal contacts and through the snowball technique in the community and in the University of Aveiro. All of them were in good health, and they have no reported history of neurological disorders, head/neck diseases, or any speech, language, or hearing difficulties. All speakers were free of upper respiratory tract infection, and were excluded (1) if they were current smokers or had smoked within the previous 5 years; and (2) if they wore hearing aids.

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.

Considering the speakers' anatomical characteristics, young (Y) and old (O) female speakers presented similar height ( $Y = 161.80 \pm 5.67$ ;  $O = 161.40 \pm 3.07$ ) but different weight ( $Y = 58.00 \pm 8.46$ ;  $O = 66.60 \pm 11.83$ ), which is reflected in the greater occurrence of

overweight among older women (Table 1). All speakers have 9 years or more of education, with the majority (90%) having completed higher education.

**Table 1.** Speakers' anatomical characteristics.

Speaker	Weight (kg)	Height (cm)	BMI *	Weight Status
Y1	72	158	28.8	Overweight
Y2	62	160	24.2	Normal weight
Y3	50	158	20.0	Normal weight
Y4	57	173	19.0	Normal weight
Y5	49	160	19.1	Normal weight
O1	48	158	19.2	Normal weight
O2	60	167	21.5	Normal weight
O3	80	162	30.5	Obesity class I
O4	67	160	26.2	Overweight
O5	78	160	30.5	Obesity class I

$$* \text{ BMI} = \frac{\text{weight(kg)}}{(\text{height(m)})^2}$$

## 2.2. Corpus

The corpus consisted of all EP oral vowels [i], [e], [ɛ], [a], [o], [ɔ], [u], [ɪ], and [ʊ] in pseudoword context and in isolated context. The pseudoword list contained 'pV.Cv sequences (started with the labial voiceless stop consonant [p] to reduce coarticulation with the following vowel), 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 the stressed position. The last vowel (i.e., v) corresponds only to the vowels [u], [ɪ], and [ʊ].

The stimuli were embedded in a carrier sentence "Em pVCv temos V" (*In pVCv we have V*), where the last V was considered an isolated vowel and corresponded to the same vowel that occurred in the stressed position of the pseudoword. For each vowel, three different pseudowords were selected. The list of 27 pseudowords used in this study is listed in Table 2 in IPA symbols; however, the pseudowords were presented to the speakers in the respective writing convention of EP (e.g., ['peke]—"pêca"). Each carrier sentence was repeated 3 times. Thus, each speaker produced 81 individual utterances (i.e., 9 repetitions of each vowel, per context) and needed approximately 30 min to complete the task.

**Table 2.** List of pseudowords per vowel (International Phonetic Alphabet).

Vowels		Pseudowords		
front	[i]	['pipi]	['pitɛ]	['pikɛ]
	[e]	['pepi]	['petu]	['peke]
	[ɛ]	['pɛpi]	['pɛtɛ]	['peku]
central	[ɪ]	['pipi]	['pitɛ]	['pikɛ]
	[ʊ]	['pepi]	['petɛ]	['peke]
	[a]	['papi]	['patɛ]	['paku]
back	[u]	['pupi]	['putu]	['puke]
	[o]	['popi]	['potu]	['poku]
	[ɔ]	['pɔpi]	['pɔtɛ]	['pɔke]

## 2.3. Data Acquisition

The experimental setup for data acquisition is shown in Figure 1. The participants were asked to seat, facing a computer screen displaying prompts, and to wear a stabilization helmet [66], in order to ensure that neither the speaker's head nor the transducer moved during the experiment.



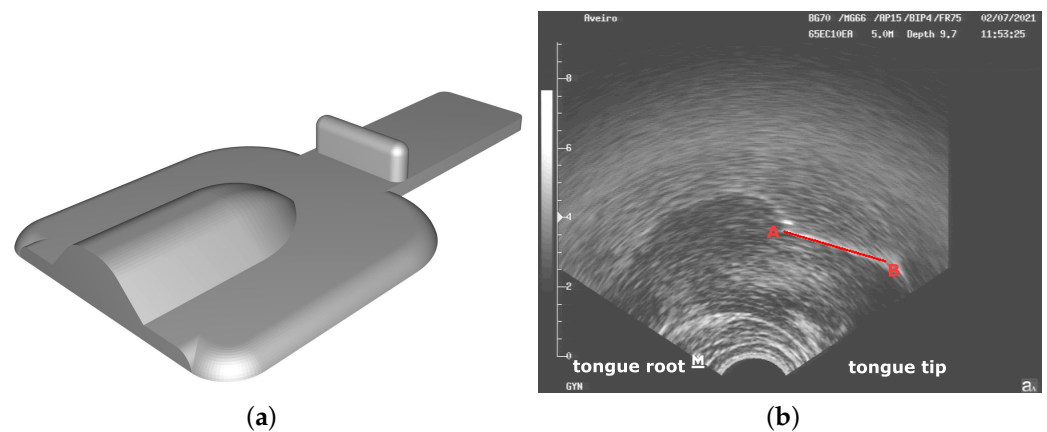


**Figure 1.** Experimental setup for recording ultrasound: the participant is seated in front of a microphone, wearing a stabilization helmet to keep the ultrasound probe positioned below the chin throughout the experiment. The prompts are presented on a computer monitor.

Synchronous acquisition of US images and speech sounds was performed through Articulate Assistant Advanced software (AAA) [67] and took place in a quiet room using an endocavitary probe (65EC10EA) with 90° field of view positioned under the participants' chin. The US images were collected at 60 frames per second using a Mindray DP6900 US equipment with depth set to 97 mm. The speech audio was collected with a Philips SBC ME400 microphone connected to an external sound system (UA-25 EX USB). The synchronization between the video and audio streams was performed in AAA based on the synchronization pulses introduced, during recording, by a SyncBrightUp module [68,69].

Instructions were provided prior to recording to ensure familiarity with the speech materials. Speakers were instructed to read the sentences at a comfortable pace. The speech material was presented in three randomized blocks (i.e., front ([i], [e], [ɛ]), central ([i], [e], [a]), and back ([u], [o], [ɔ]) vowels). At the start of each block, a recording of the production of the sequence [tatatata] was obtained to provide additional data for further assessing the audio–video synchronization during the data processing stage. In addition, each block began and finished with the production of the sustained vowel [a] and the recording of the bite plane. The bite plane was recorded in order to image the occlusal plane, which is a reliable method for the definition of horizontal and vertical orientations in the vocal tract [52,70]. That is, the speaker was asked to bite and press their tongue against a flat plastic plate, with 4 cm length, back from the upper incisor, which results in their tongue bulging upward at the back edge of the bite plate [52,70]. This bite plate was designed by the authors for this experiment and is shorter than the one used in Scobbie et al. [52] to prevent the gag reflex (see Figure 2a). The bite plate was 3D printed in PETG filament (natural color).

The probe orientation was adjusted between the shadows of the mandible and hyoid bones, taking into account the different anatomical characteristics of each speaker. Then, the bite plane sequences and the contours of the sustained vowels [a] were used to obtain the referential for each speaker. To ensure that for each speaker, there is at least one bite plane trace to obtain the referential, the probe orientation was fixed along the session.



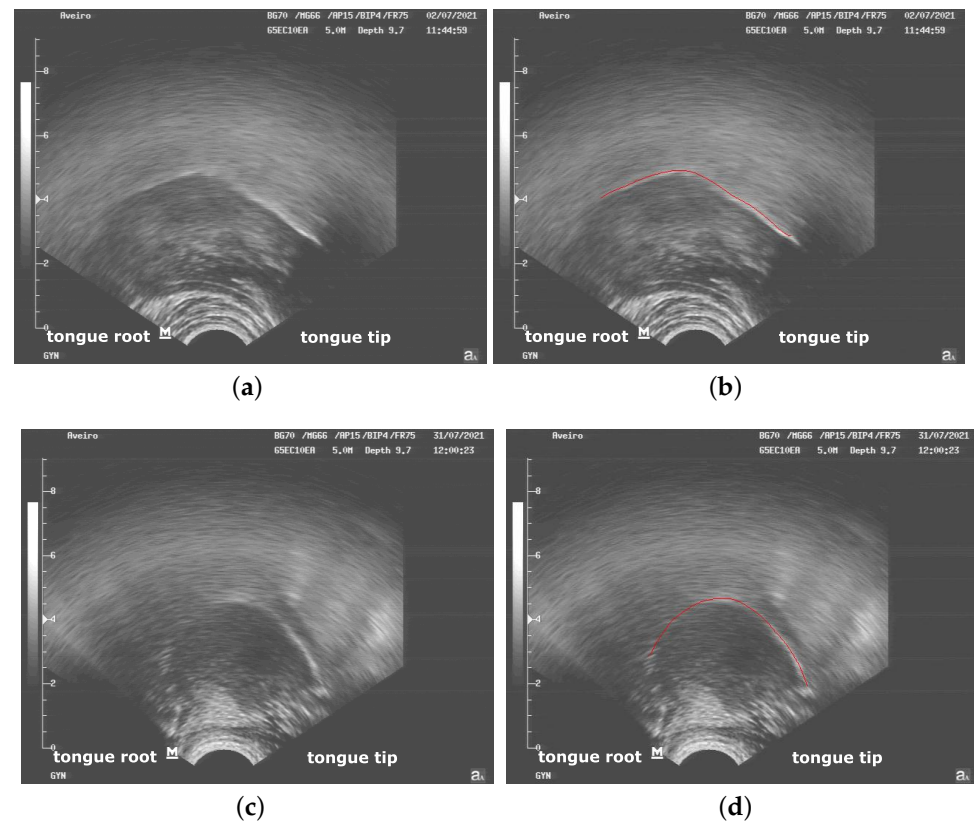
**Figure 2.** To ensure a common referential across blocks and sessions, US acquisitions were considered, including a bite plate (a). The corresponding images were annotated (b), and the data informed contour adjustment (e.g., rotation) before analysis. (a) Bite plate; (b) Bite plane trace.

#### 2.4. Data Processing

After acquisition, the data obtained were processed in several steps: (1) speech audio and US synchronization; (2) segmentation of the acoustic data signal and manual revision; (3) US image processing using a U-net network [71]; (4) automatic extraction of the tongue contours in Cartesian coordinates from the vowel midpoint; (5) manual verification of the extracted contours; (6) bite plane extraction for each block of vowels; (7) tongue countour rotation to ensure a common referential across blocks; (8) inspection of the US data with a particular emphasis on the [a] sustained vowels as a quality check of the previous steps; (9) export of US data contours; and (10) data pruning of the tongue contours. Each of these steps will be detailed below.

At the end of each session the ultrasound data were synchronized with the speech data by aligning the audio tone, which was triggered at the onset of each recording, and the video flash (i.e., a bright flash on the corner of the ultrasound image) [69,72]. After that, the acoustic files were exported from AAA software in WAV format for automatic segmentation at the word and phoneme level using WebMAUS General for Portuguese language (PT) [73]. Then, the accuracy of the target vowel boundaries were manually checked in Praat software [74] and corrected if necessary. A total of 22 recordings were discarded (approximately 1.4% of trials) due to problems with the recordings (e.g., the participants misread the target word or the target vowel). After acoustic segmentation, all frames that corresponded to the target vowels were extracted from the US videos using a Python script. A database was devised, containing both vowel contexts, pVCv, and isolated vowels.

The US images were processed, and splines were automatically fitted to the surface of the tongue across the vowel's labeled duration using a U-Net adapted from Ronneberger et al. [71] (already used in US images by [75,76]). This encoder–decoder model was trained with a dataset of 7765 US images of different speakers that were manually annotated by four trained analyzers. In order to virtually increase the amount of training data, data augmentation was also used to train the neural network, and the final training set consisted of 9131 images. Figure 3b,d show two examples of the automatic tongue contour obtained by the developed method.



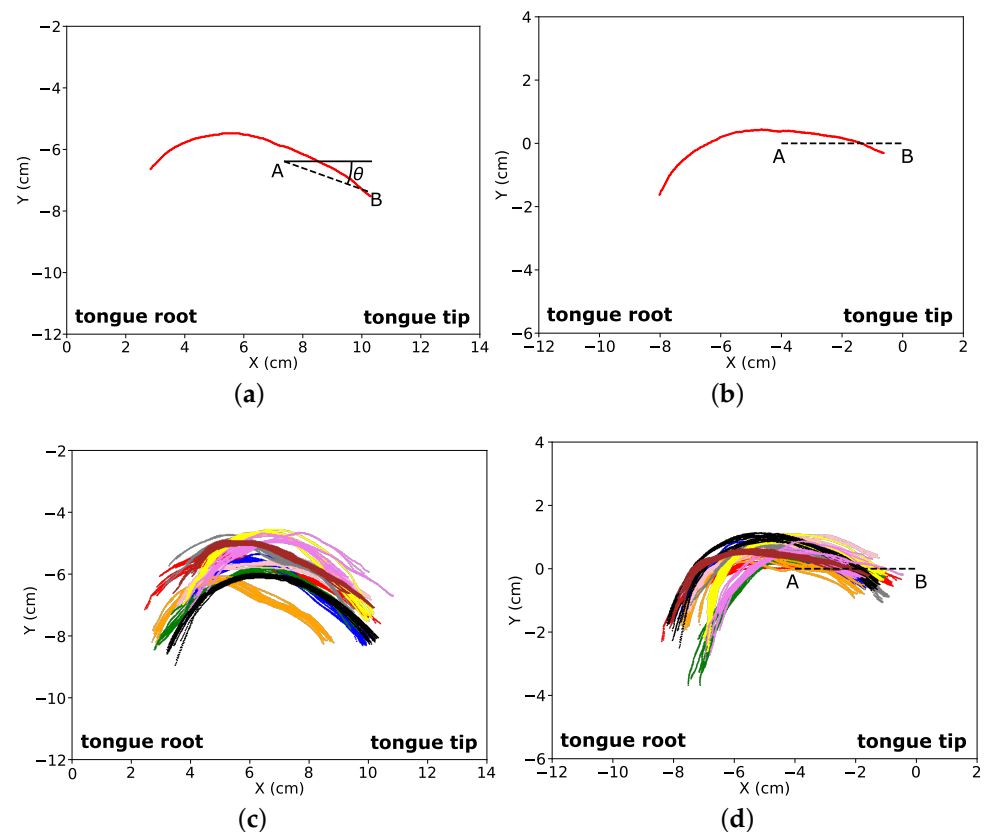
**Figure 3.** Illustrative US images with the automatic tongue spline of the tongue contour for vowel [a] (a,b) and vowel [u] (c,d). Left: raw US images; Right: US images with automatic tongue spline.

For the present study, only the tongue contour corresponding to the temporal mid-point of the vowels was exported, which consistently contained an articulatorily steady part of the vowel [70]. When the vowels presented an even number of frames, both frames were exported. The output data structure contains the Cartesian coordinates of the automatic tongue contour of each frame, as a list of points (x,y values in pixel). Given the challenging nature of the images, and to ensure the reliability of the data, the automatic tongue segmentation of the central frames of the vowel occurrences was revised by three annotators with experience in speech production analysis.

In total, 2303 splines were obtained through automatic tongue segmentation, and 227 images were excluded due to the undefined border of the tongue or extremely dark US image. The final number of splines, after manual check, were 2076, which corresponded to 1440 vowels.

The bite plane was extracted for each block of vowels. The bite plane recording that appeared to have the most stable pressing of the tongue was used, and this recording may have been at the beginning or at the end of each randomized block. Each bite plane is represented by the points A and B (see Figure 2b). For each point A and B, the coordinates x and y in pixels were determined to obtain the bite plane line. Nonetheless, for Speaker O5, it was not possible to get one bite plane per block. As the probe orientation was fixed along the session, the tongue contours were rotated based on the available speaker's bite planes. Then,  $\theta$  values were calculated as the angle between the bite plane and the horizontalization of the occlusal plane (see Figure 4a).





**Figure 4.** Example of unrotated (a) and rotated (b) tongue tracing of one sustained vowel [a] (top). Tongue contour of each sustained vowel [a] of all speakers unrotated (c) and rotated (d) (bottom). The dashed line represents the annotation of the bite plate.

In order to standardize the rotation of the data across the speakers, for each block of vowels, the tongue contour data were rotated based on the occlusal plane obtained in the corresponding block [52,54,70]. As seen in Figure 4, the occlusal plane was observed to be parallel to the  $x$  axis after rotation (see Figure 4b). During rotation, pixel to cm conversion was made. Data were exported after rotation, and the origin of the coordinates system corresponds to the upper incisors, which are 4 cm from the back of the bite plate.

Four frames of each sustained vowel [a] were exported and manually verified. The most complete contour of each [a] was chosen, and the contours were overlapped to inspect the US data. Figure 4 presents unrotated (Figure 4c) and rotated (Figure 4d) [a] contours of all speakers. Despite the anatomical differences between speakers, the raw coordinate values are partly comparable across the speakers, since the tongue advancement/retraction consists in a distance to the upper incisors. As seen in Figure 4, after rotation, the tongue contours of the sustained vowels [a] of all speakers tend to be more overlapping.

The tongue contours selected for analysis were revised to detect erroneous/incomplete segmentations mostly due to poor image quality rendering the tongue incomplete. This resulted in the exclusion of 113 contours. Additionally, poor US image quality for some speakers (namely for O2 and O5), vowels (mostly vowel [i]), and/or context (mostly in pVtv context) resulted in less data usable for analysis. Table 3 presents the number of tongue contours that are considered for study after the manual check of the US data.

Overall, as expected, the best US images come from sounds where the tongue surfaces are fairly flat and gently curved, such as the central vowels ([a] and [e]). Generally, a less discernible tongue contour occurs at the tip and back regions of the tongue, and vowels that have steep slopes, such as [i] or [u], tend to present worse images. Stone [50] also explained that the edges perpendicular to the beam will image best, and edges more than

50 degrees from perpendicular begin to image poorly. Other studies of vowel production report similar difficulties [62,63].

**Table 3.** Number of vowels analyzed per speaker and vowel type after removing some of the data due to poor image quality.

	Speaker										Total
	Y1	Y2	Y3	Y4	Y5	O1	O2	O3	O4	O5	
[i]	17	18	14	17	17	2	5	6	1	0	97
[e]	18	18	17	18	18	17	10	16	8	2	142
[ɛ]	18	18	18	18	18	17	16	15	16	7	161
[i]	17	18	18	18	17	17	10	17	13	6	151
[ɐ]	18	18	18	18	18	16	18	18	16	18	176
[a]	18	18	18	17	16	18	14	18	17	17	171
[u]	13	18	18	17	18	18	0	14	12	14	142
[o]	18	16	14	18	18	18	0	11	12	17	142
[ɔ]	18	15	18	18	18	18	0	16	17	17	155
<b>Total</b>	<b>155</b>	<b>157</b>	<b>153</b>	<b>159</b>	<b>158</b>	<b>141</b>	<b>73</b>	<b>131</b>	<b>112</b>	<b>98</b>	<b>1337</b>

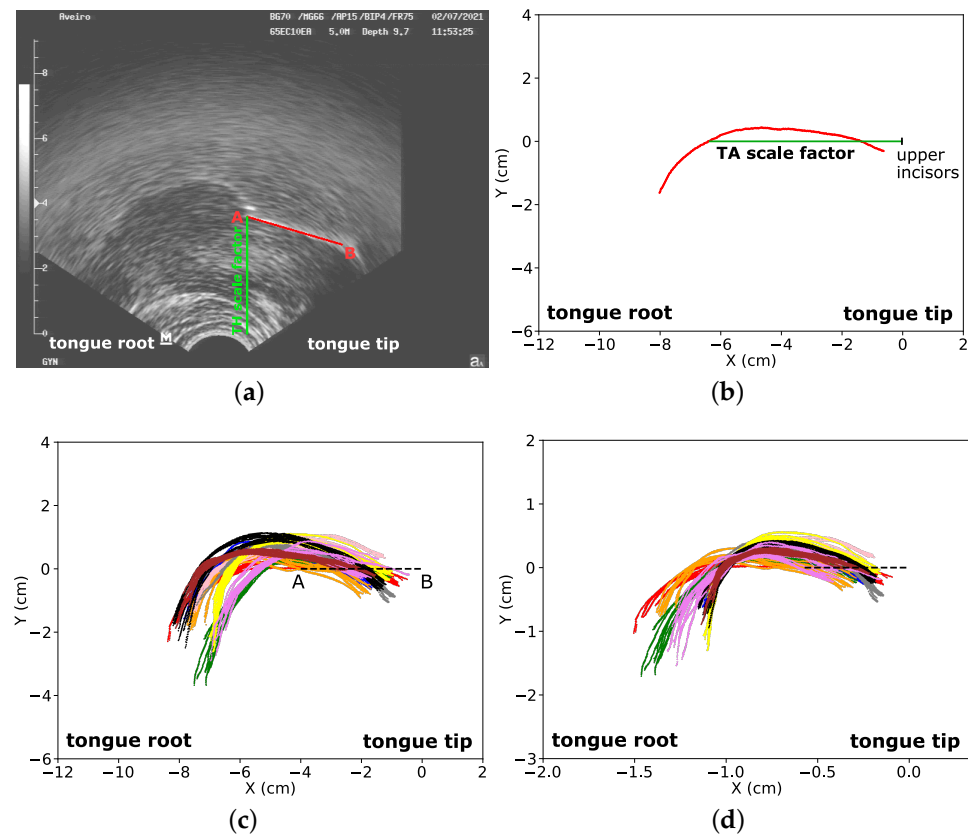
## 2.5. Articulatory Measures and Normalization

For each vowel token, two parameters were extracted, namely the y-coordinate of the highest point of the tongue's contour (i.e., tongue height, TH) and the corresponding x-coordinate, which reflects the front back position of the tongue (i.e., tongue advancement, TA). Thus, the TH corresponds to the distance between the bite plane and the highest point of the contour, while TA was measured as the distance to the upper incisors. The TA values are always negative; the higher values (in module) indicate greater tongue retraction and lower values (in module) suggest greater tongue advancement of the highest point of the tongue. When the vowels present two central frames, the mean values of TA and TH were considered.

In order to compare average tongue measures between age groups, it was necessary to adopt an approach toward a normalization of the data based on vocal tract size and US probe orientation. Therefore, the raw measures of vowels are sensitive to anatomical differences between speakers, with expected larger vowel spaces for larger speakers [54]. One potential normalization method would consist of z-score normalization, but in this case, a full set of vowels or a normalization as a function of the distance to a reference vowel would be required [54]. On the other hand, Zharkova et al. [51] applied a normalization method based on the assumption that the length of tongue contour correlates with the overall size of the vocal tract, and the proportion of tongue imaged in the ultrasound display is similar across speakers.

Nonetheless, in this study, the normalization is independent of the proportion of the tongue imaged for each speaker, but it involves the assumption that each distance used as scale factor (TA and TH scale factor) correlates with the overall size of the vocal tract. For TH, the normalization for vocal tract height was carried out on the basis of the distance of the back of the bite plate to the virtual origin of the probe (in mm) for each speaker by block of vowels (see Figure 5a). The normalization factor for vocal tract length was based on the tongue contour of the sustained vowels [a], because the tongue tip and the larynx are low during the production of this vowel; thus, the imaged tongue contour tends to be better [51]. The TA normalization factor was based on the distance, in mm, of the upper incisors to the interception between the occlusal plane and the tongue root surface in the production of sustained [a] (mean of all [a] productions), for each speaker (see Figure 5b).

For each speaker, TH and TA normalization was applied, multiplying each raw value of TH and TA by the corresponding normalization factor. Figure 5c,d illustrate the normalization procedures.



**Figure 5.** Top: Illustration of normalization measures for TH (a) and TA (b). Bottom: Sustained vowels [a] contours of all speakers rotated (c) and scaled (d).

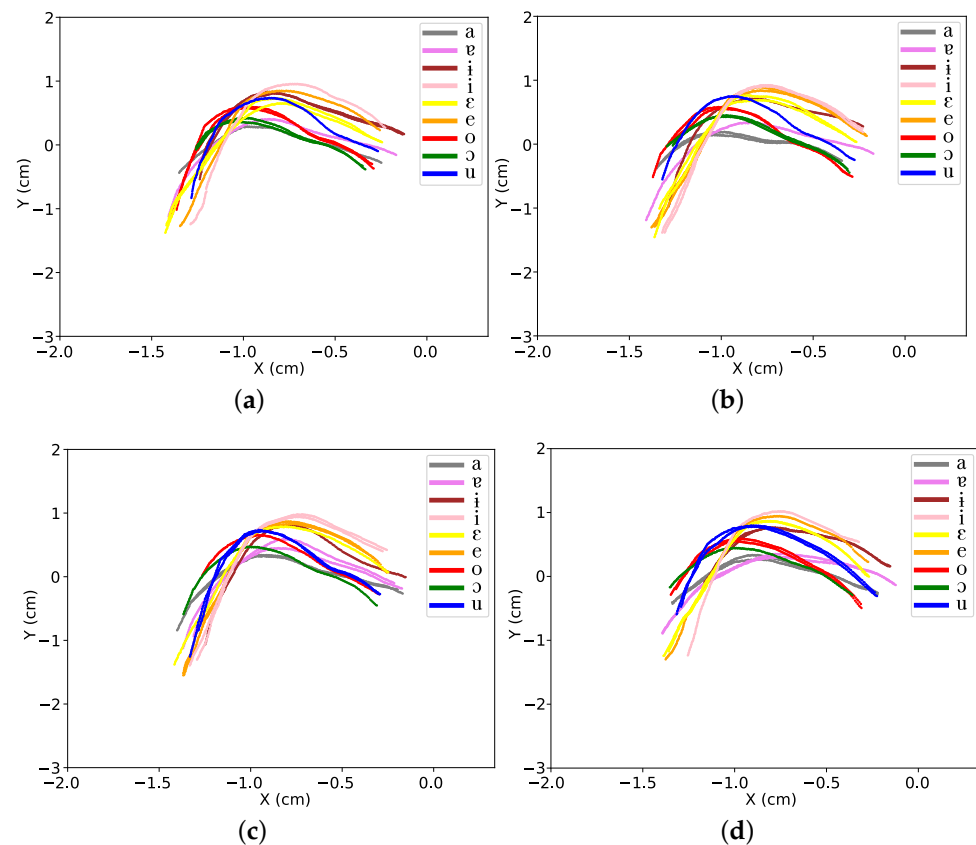
### 3. Results

In this section, a summary of the main findings on EP vowel tongue configurations by age group is presented. In addition, some considerations about vowel context effect, and inter and intra-speaker differences are reported. Note that for some older speakers (see Table 3), there are no tongue contours available for some vowels, which does not allow a more comprehensive comparison between age groups to be carried out. In this section, the scaled US data are used (i.e., TA and TH obtained after normalization). The analysis starts with a brief inspection of the tongue contours. Thus, some examples of tongue contours for all vowels by context are presented in Figure 6 for one speaker (speaker Y1).

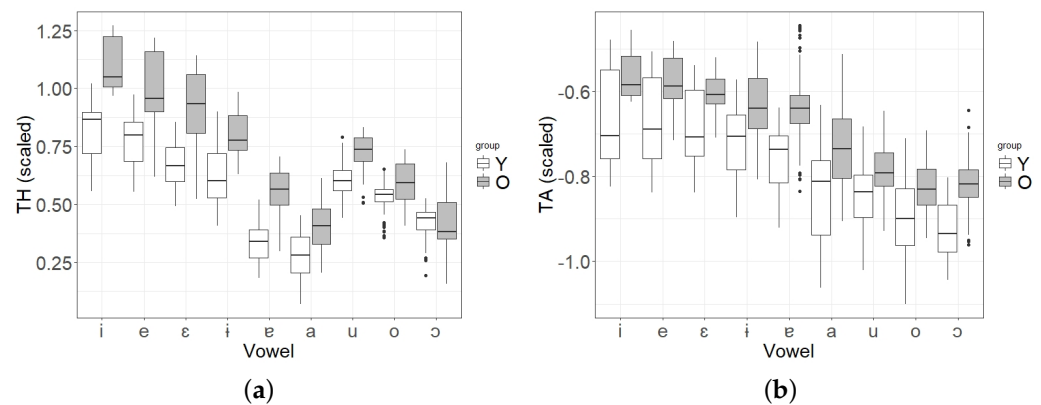
Considering the articulatory measures, Figure 7 summarizes the TH (Figure 7a) and TA (Figure 7b) values obtained by vowel for each age group. In this first analysis, both contexts (vowels in pVCv sequence and isolated) are considered. In general, concerning TH, the older females showed higher values, except for vowel [ɔ]. In general, the front vowels presented higher TH values, and the vowels [a] and [ɔ] presented the lower values, mainly for the older females. For young females, the central vowels [e] and [a] showed similar TH, while for older females, the vowel [e] presented higher TH values than vowel [a]. In addition, old females presented identical values of TH for vowel [a] and [ɔ].

Regarding TA, the lowest values (in module) were obtained for the old group. For both groups, the front vowels tend to present the lowest TA, and the vowels [o] and [ɔ] revealed the highest TA values (in module). Variability tends to be higher in TA in comparison with TH values, mainly for young females (i.e., the variability of TA tends to decrease with age).

To complement this initial analysis, the individual variability is also explored through the articulatory vowel space and the vowel articulatory cluster size by speaker and vowel context.



**Figure 6.** Example of one tongue contour by vowel for each context for speaker Y1. (a) pVpv context; (b) pVtv context; (c) pVkv context; (d) isolated context.



**Figure 7.** Boxplots of TH (a) and TA (b) values by vowel and age group.

Figure 8 represents the articulatory vowel space defined by scaled TA and TH values of the cardinal EP oral vowels ([a], [i], and [u]) for each speaker and vowel context (i.e., pVCv and isolated vowel). Each vowel is represented by the mean TH and TA. As four old females (O1, O2, O4, and O5) presented incomplete data, it was not possible to define the articulatory space of all speakers by vowel context. Furthermore, the young females also showed great variability in the shape of the articulatory vowel space. Concerning the vowel context, it can be observed that articulatory space tends to be smaller when vowels occur in pVCv sequences comparing with isolated vowels, namely for the speakers Y2, Y3, Y5, and O3.

Since our main goal is the study of age effects, in Figure 9, the articulatory vowel space of the two age groups is compared separately for each production context. The articulatory space was defined by the mean of the TA and TH values of the cardinal EP oral vowels for each speaker by vowel context.

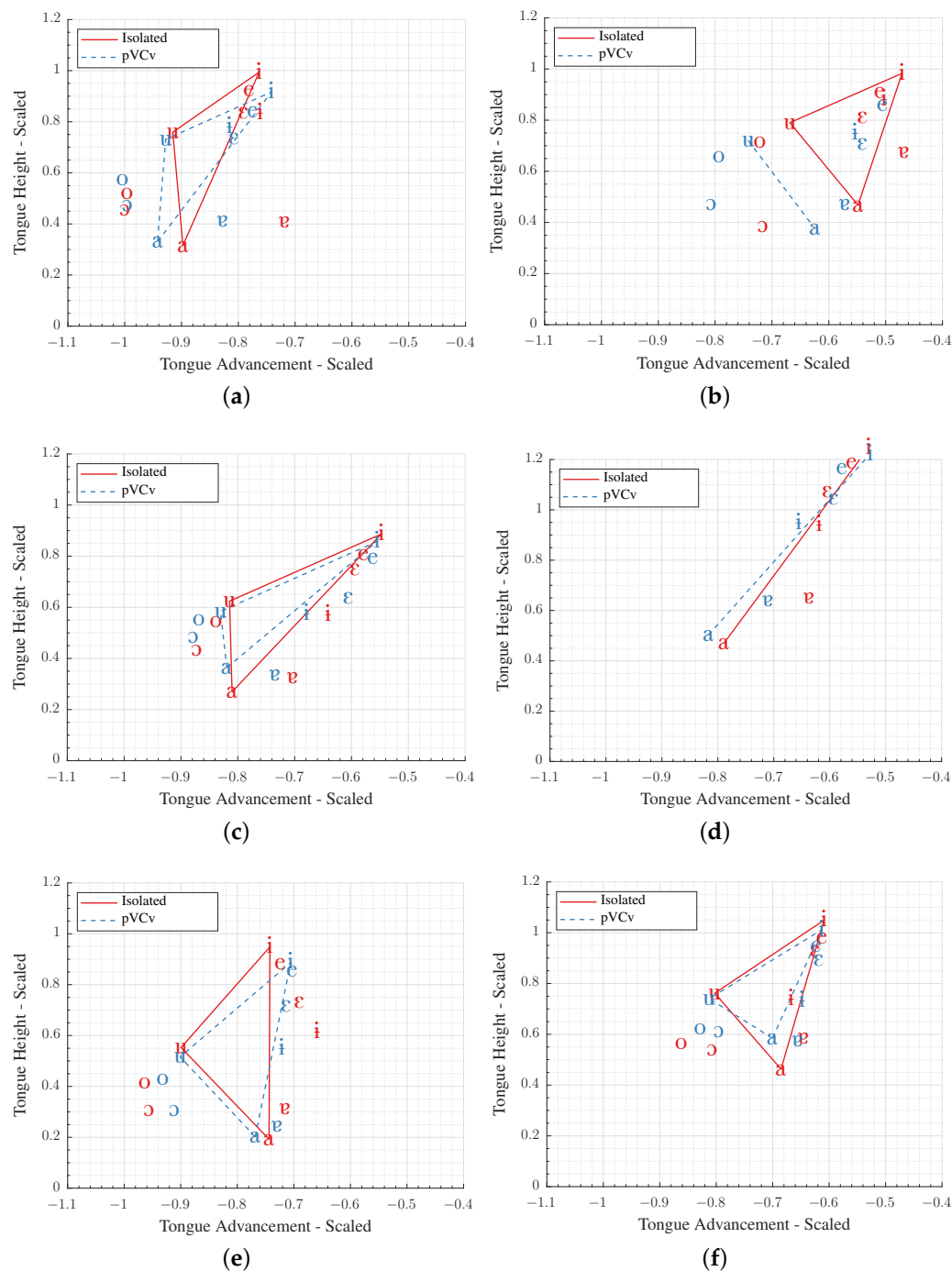
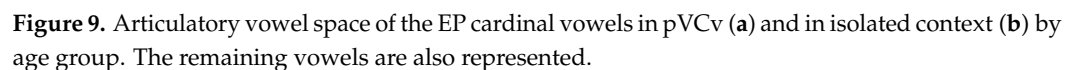
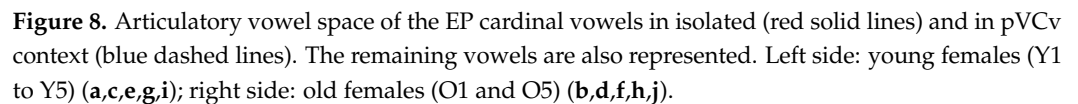


Figure 8. Cont.

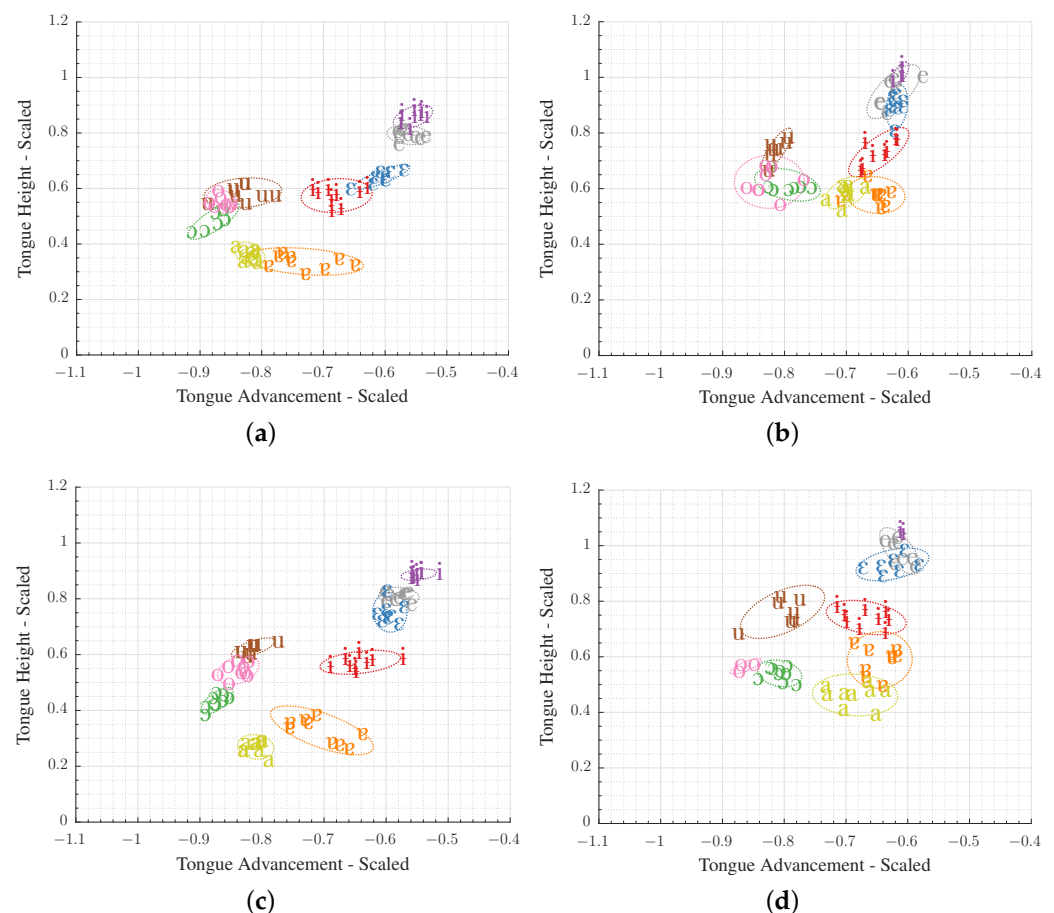




The vowel articulatory space tends to be different, in shape and size, between the younger and older females in both contexts. Old females produced vowels with higher TH and a notable decrease (in module) of the TA (i.e., the highest point of the tongue contour tended to be more advanced). Consequently, these results are in line with some of the observations made above. The vowel articulatory space differences observed between both contexts is more pronounced for the young females. That is, the articulatory vowel

space area, namely for pVCv sequences, tended to be smaller in the young females. In old females, the differences between contexts were smaller, but a tendency to TA decrease (in module) for isolated vowels comparing with vowels produced in pVCv sequences (i.e., isolated vowels tend to be more advanced) was observed.

The plots in Figure 8 only provide the average, but variability of productions is also very important to analyze. Figure 10 represents individual productions (i.e., the TH and TA of the total number of occurrences of all EP oral vowels) and information regarding dispersion based on ellipses for two speakers (one young and one old) by vowel context. The ellipses surround values that fall within  $2\sigma$  of the mean [24]. In those graphs, a considerable variability between speakers can be observed.



**Figure 10.** Vowel articulatory cluster size of all EP vowels of two speakers in pVCv sequences (top) (a,b) and in isolated context (bottom) (c,d). Left side: young female (Y2); Right side: old female (O3).

#### 4. Discussion

This study contributes to increasing knowledge on EP aging speech, providing an articulatory perspective of the effects of age in all oral vowels of the EP based on articulatory measures: TH and TA. The present study extends in many ways our previous pilot research by reporting articulatory data from more female speakers and by ensuring an inter-speaker comparison through the application of normalization procedures. In addition, the automatic method of contours tracing was improved.

In general, the results of this articulatory study reveal that the highest point of the tongue tends to be higher for the older females and more advanced compared to the younger females (except the TH of the vowel [ɔ]). The increase of the TH is in accordance with our previous acoustic study [16] of aging speech for EP. In other words, the tongue raising (higher TH) is correlated with a decrease in F1 [9,56], which has been reported for

both EP [15,16] and other languages [4,5,44,77,78] with aging. However, Pellegrini et al. [17] reported higher F1 values for vowels produced by older EP females.

Regarding TA, the results indicate a tendency to tongue advancement in the older group, which is correlated with an increase in F2 [9,56]. The tongue advancement observed for older females is not in line with our previous acoustic study [16], where a tendency for F2 decrease with aging occurred. Yet, some studies have reported vowel-specific changes with age and gender [5,11,17,78,79]. For EP, Pellegrini et al. [17] observed a trend for F2 decrease in back vowels and an increase in front vowels. Nonetheless, Sonies et al. [23] reported a reduction in tongue retraction on the vowel [a] in older adults, and they suggested that older speakers tend to hold the tongue in a more anterior position in the mouth than younger speakers. The authors [23] attributed this reduction in tongue retraction to a normal compensatory tongue motion, which is apparently due to alterations in the function of the suspensory muscles of the tongue [80].

Beyond the observed displacement of the articulatory vowel space with age (i.e., more advanced) (see Figure 9), the vowels' space tends to be more reduced in younger than in older females. For EP, Pellegrini et al. [17] also reported an acoustic space that is more reduced for younger females. Although formant frequencies can be affected by other articulatory adjustments than the tongue movements, it would be interesting to study the acoustic–articulatory relationship of the data. Furthermore, Xue and Hao [44] suggested that the lowering of formants with age might be due to the lengthening of the vocal tract with aging, but in the present study, a tongue raising and a tongue advancement were observed for the older females.

Additionally, the articulatory measures show great variability, as it can be observed by the different format of the individual articulatory vowel spaces and the vowel articulatory clusters. The comparison of the amount of variability within and across speakers in both age groups needs to take into account the fact that older speakers present less data, which does not allow drawing solid conclusions. In addition, there does not appear to be a direct relationship between speaker size (BMI) and the length of the tongue contour or the size of the acoustic triangle obtained.

Concerning the vowel type, our results indicate a significant difference in tongue height position between the EP oral vowels examined, mainly between front and back vowels. As expected, for each block of vowels, the TH is higher for close vowels and lower for open vowels. An interesting observation is that for young females, the central vowels [ɐ] and [a] show lower values and similar TH among each other, while for older females, the vowel [ɐ] presents higher TH than vowel [a]. Regarding TA, front vowels tend to present a more advanced higher point of the tongue contour.

Regarding the vowel context, as in the pilot study [22], the articulatory space tends to be smaller when vowels occur in pVCv sequences comparing with isolated vowels, mainly for younger females. The vowel articulatory space reduction observed for vowels in pVCv sequences in comparison with isolated vowels might be related with the tendency to hyperarticulate isolated vowels. This type of effect was also evident between vowels in clear speech versus in conversational speech [63], or in long vowels versus short vowels [81], for other languages.

While for young females, a reduction in the articulatory vowel space between contexts was observed, and for older females, the vowel space tends to be more advanced for isolated vowels. This different pattern between vowel contexts with age might be related to specific articulatory adjustments of the older females.

In this study, the extraction of the tongue contours (using a U-Net adapted from Ronneberger et al. [71]) and the determination of the highest position of the tongue body for young and old Portuguese females were done automatically. Figure A1 shows the TA and TH scaled measures obtained of all tongue contours without any manual verification. Compared with the data obtained after manual inspection of the tongue contours (see Figure 7), similar tendencies for each vowel by age group are observed, even though a higher amount of outliers occur. In a broader perspective, this automatic method may

enable an unsupervised extraction and analysis of these tongue measures, making possible the analysis of large amounts of US data for young and old speakers. Notwithstanding, more methodological work is necessary to improve the automatic tongue tracing to deal with extraction errors, and the accuracy of this method still needs to be measured in order to avoid the manual revision of the tongue contours, as this is a time-consuming task.

As the dataset is small, it cannot be excluded that speaker-specific strategies interact with group behavior [42]. In addition, the US images could be affected by many factors, such as the anatomical differences (e.g., facial profile, thickness of the fat tissue around the neck), or the pressure or position of the transducer between speakers [43,54]. Furthermore, the noisy nature of the images make the segmentation demanding, and several unclear images had to be excluded from the study.

Moreover, the analysis of the TH and its horizontal location (TA) can be inaccurate when the highest point of the tongue is not located at the narrowest point of the tongue–palate constriction [62]. However, as the palate traces were largely unreliable for the majority of the speakers, it is not possible to implement a measure of constriction degree for the current data. Thus, reducing the lingual configuration to a single point is a convenient methodological solution, but it is far from adequate in giving a comprehensive description of the lingual articulation [62].

As raw distance measures between vowels are sensitive to anatomical differences between speakers, this study proposes a new normalization approach that is independent of the proportion of the tongue imaged for each speaker. However, this normalization method requires further validation.

Considering that the present study is a preliminary analysis of articulatory changes in lifetime, the researchers intend to extend the study to the male gender and to analyze other measures such as the tongue root and the total vowel contour through Smoothing Spline ANOVA [63,82,83]. In addition, static and dynamic studies of the vowel tongue measures need to be addressed to investigate articulatory movement and velocity.

## 5. Conclusions

In this study, the age-related articulatory changes are investigated through the analysis of the US tongue contours of nine EP oral vowels produced by 10 female speakers aged between 23 and 73. Advancing our preliminary work on this topic, an improved automatic method for tongue contour tracing allowed expanding the number of considered speakers and supported a novel normalization procedure for inter-speaker comparison.

In light of this new articulatory data, it can be concluded that the tongue tends to be higher and more advanced with aging for almost all vowels, meaning that the vowel articulatory space tends to be higher, advanced, and bigger in older females. Concerning the context, the vowel space tends to be more advanced for isolated vowels comparing with vowels produced in disyllabic sequences for older females, while younger females tend to present a sharp reduction in the articulatory vowel space in disyllabic sequences.

These results contribute to an accurate articulatory description of EP oral vowels, and they also provide valuable insights about the vowel articulatory normal patterns of aging among female Portuguese adults. As this study provides information from adults who speak a language different from English, it might help to better understand cross-linguistic similarities and language-particular features of vowel aging. Furthermore, these data are important as a reference for the clinical assessment and treatment of different speech disorders, which are often age-related, and to provide articulatory information for speech synthesis based on production models.

In the long term, the systematic approach enabled by the adopted automatic method of tongue contour tracing, along with the normalization supporting inter-speaker comparison, may open the way toward broader lifespan articulatory studies and the development of larger databases with speakers from infancy to old age.

**Author Contributions:** Conceptualization, L.A., A.R.V., C.O. and A.T.; methodology, L.A., A.R.V., F.B., A.T., S.S., P.M. and C.O.; validation, L.A., A.R.V., F.B., A.T., C.O. and S.S.; formal analysis, L.A., A.R.V., F.B. and A.T.; investigation, L.A. and A.R.V.; software, L.A., F.B. and A.T.; data curation, L.A., F.B. and A.T.; writing—original draft preparation, L.A. and A.R.V.; writing—review and editing, F.B., A.T., S.S., P.M. and C.O.; visualization, L.A. and F.B.; supervision, A.T., S.S. and C.O.; project administration, A.T., S.S. and C.O.; funding acquisition, L.A., A.T., S.S. and C.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** 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). The first author was funded by the grants SFRH/BD/115381/2016 and COVID/BD/151744/2021 (funded by FCT, through FSE and by CENTRO2020).

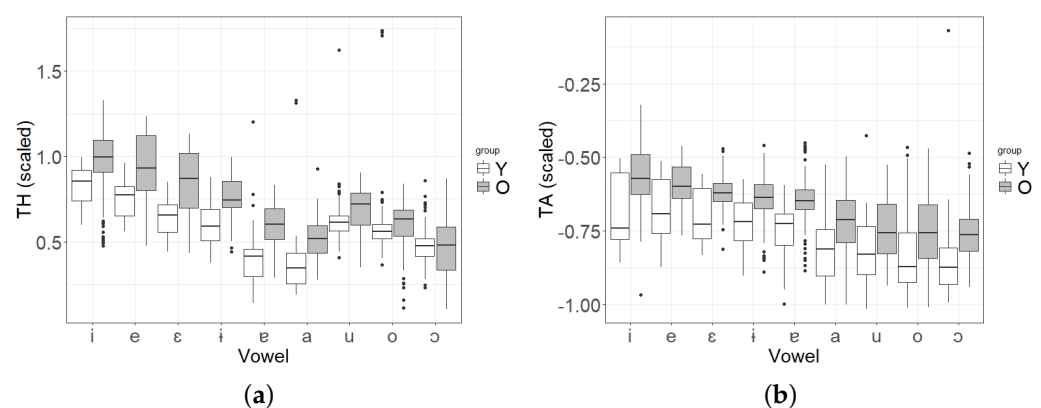
**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Escola Superior de Enfermagem de Coimbra, Portugal (approval number 639/12-2019).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Acknowledgments:** We are very grateful to all the adults who contributed as speakers.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. TA and TH Scaled Measures Obtained of All Tongue Contours without Any Manual Verification



**Figure A1.** Boxplots of TH (a) and TA (b) scaled values by vowel and age group of all tongue contours automatically obtained.

## References

1. Hermes, A.; Mertens, J.; Mücke, D. Age-Related Effects on Sensorimotor Control of Speech Production. In Proceedings of the Interspeech, Hyderabad, India, 2–6 September 2018; pp. 1526–1530.
2. Makiyama, K.; Hirano, S. *Aging Voice*; Springer: Singapore, 2017.
3. Massimo, P.; Elisa, P. Age and Rhythmic Variations: A Study on Italian. In Proceedings of the Interspeech, Singapore, 14–18 September 2014; pp. 1234–1237.
4. Linville, S.E. *Vocal Aging*; Singular Thomson Learning: San Diego, CA, USA, 2001.
5. Schötz, S. Perception, Analysis and Synthesis of Speaker Age. In *Linguistics and Phonetics*; Lund University: Lund, Sweden, 2006; Volume 47.
6. Goozée, J.; Stephenson, D.; Murdoch, B.; Darnell, R.; Lapointe, L. Lingual kinematic strategies used to increase speech rate: Comparison between younger and older adults. *Clin. Linguist. Phon.* **2005**, *19*, 319–334. [\[CrossRef\]](#)
7. Vipperl, R.; Renals, S.; Frankel, J. Ageing voices: The effect of changes in voice parameters on ASR performance. *Eurasip Aud. Speech Music. Process* **2010**, *2010*, 525783. [\[CrossRef\]](#)
8. McDougall, K.; Nolan, F. Discrimination of speakers using the formant dynamics of /u:/ in British English. In Proceedings of the International Congress of Phonetic Sciences (ICPhS XVI), Saarbrücken, Germany, 6–10 August 2007; pp. 1825–1828.



9. Kent, R.D.; Vorperian, H.K. Static Measurements of Vowel Formant Frequencies and Bandwidths: A Review. *J. Commun. Disord.* **2018**, *74*, 74–97. [[CrossRef](#)] [[PubMed](#)]
10. Fant, G. *Acoustic Theory of Speech Production: With Calculations Based on X-ray Studies of Russian Articulations*, 2nd ed.; Walter de Gruyter: Berlin, Germany, 1970.
11. Eichhorn, J.T.; Kent, R.D.; Austin, D.; Vorperian, H.K. Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women. *J. Voice* **2018**, *32*, 644.e1–644.e9. [[CrossRef](#)] [[PubMed](#)]
12. Das, B.; Mandal, S.; Mitra, P.; Basu, A. Effect of aging on speech features and phoneme recognition: A study on Bengali voicing vowels. *Int. J. Speech Technol.* **2013**, *16*, 19–31. [[CrossRef](#)]
13. Tykalova, T.; Skrabal, D.; Boril, T.; Cmejla, R.; Volin, J.; Rusz, J. Effect of Ageing on Acoustic Characteristics of Voice Pitch and Formants in Czech Vowels. *J. Voice* **2020**, *35*, 931.e21–931.e33. [[CrossRef](#)]
14. Rastatter, M.P.; McGuire, R.A.; Kalinowski, J.; Stuart, A. Formant frequency characteristics of elderly speakers in contextual speech. *Folia Phoniatr. Logop.* **1997**, *49*, 1–8. [[CrossRef](#)]
15. Albuquerque, L.; Oliveira, C.; Teixeira, A.; Sa-Couto, P.; Freitas, J.; Dias, M.S.M. Impact of Age in the Production of European Portuguese Vowels. In Proceedings of the Interspeech, Singapore, 14–18 September 2014; pp. 940–944.
16. Albuquerque, L.; Oliveira, C.; Teixeira, A.; Sa-Couto, P.; Figueiredo, D. A comprehensive analysis of age and gender effects in European Portuguese oral vowels. *J. Voice* **2020**. [[CrossRef](#)]
17. Pellegrini, T.; Hämäläinen, A.; de Mareüil, P.B.; Tjalve, M.; Trancoso, I.; Candeias, S.; Dias, M.S.; Braga, D. A Corpus-Based Study of Elderly and Young Speakers of European Portuguese: Acoustic Correlates and Their Impact on Speech Recognition Performance. In Proceedings of the Interspeech, Lyon, France, 25–29 August 2013; pp. 852–856.
18. Albuquerque, L.; Oliveira, C.; Teixeira, A.; Sa-Couto, P.; Figueiredo, D. Age-Related Changes in European Portuguese Vowel Acoustics. In Proceedings of the Interspeech, Graz, Austria, 15–19 September 2019; pp. 3965–3969.
19. De Decker, P.; Mackenzie, S. Tracking the phonological status of /l/ in Newfoundland English: Experiments in articulation and acoustics. *J. Acoust. Soc. Am.* **2017**, *142*, 350–362. [[CrossRef](#)]
20. Belmont, A.J. Anticipatory Coarticulation and Stability of Speech in Typically Fluent Speakers and People Who Stutter Across the Lifespan: An Ultrasound Study. Master's Thesis, University of South Florida, Fowler Ave, FL, USA, 2015.
21. Neel, A.T.; Palmer, P.M. Is Tongue Strength an Important Influence on Rate of Articulation in Diadochokinetic and Reading Tasks? *JSLHR* **2012**, *55*, 235–246. [[CrossRef](#)]
22. Albuquerque, L.; Valente, A.R.; Barros, F.; Teixeira, A.; Silva, S.; Martins, P.; Oliveira, C. *The Age Effects on EP Vowel Production: An Ultrasound Pilot Study*; IberSPEECH 2021; ISCA: Valladolid, Spain, 2021; pp. 245–249. [[CrossRef](#)]
23. Sonies, B.C.; Baum, B.J.; Shawker, T.H. Tongue motion in elderly adults: Initial in situ observations. *J. Gerontol.* **1984**, *39*, 279–283. [[CrossRef](#)] [[PubMed](#)]
24. Escudero, P.; Boersma, P.; Rauber, A.S.; Bion, R.A.H. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *J. Acoust. Soc. Am.* **2009**, *126*, 1379–1393. [[CrossRef](#)] [[PubMed](#)]
25. Oliveira, C.; Cunha, M.M.; Silva, S.; Teixeira, A.; Sa-Couto, P. Acoustic Analysis of European Portuguese Oral Vowels Produced by Children. In *Advances in Speech and Language Technologies for Iberian Languages*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 328, pp. 129–138.
26. Oliveira, C.; Martins, P.; Silva, S.; Teixeira, A. An MRI study of the oral articulation of European Portuguese nasal vowels. In Proceedings of the ISCA's 13th Annual Conference, Portland, OR, USA, 9–13 September 2012; pp. 2690–2693.
27. Cunha, C.; Silva, S.; Teixeira, A.; Oliveira, C.; Martins, P.; Joseph, A.; Frahm, J. On the Role of Oral Configurations in European Portuguese Nasal Vowels. In Proceedings of the Interspeech, Graz, Austria, 15–19 September 2019; pp. 3332–3336. [[CrossRef](#)]
28. Oliveira, C.; Martins, P.; Teixeira, A. Speech Rate Effects on European Portuguese Nasal Vowels. In Proceedings of the Interspeech, Brighton, UK, 6–10 September 2009; pp. 480–483.
29. Sataloff, R.T.; Kost, K.M.; Linville, S.E. The Effects of Age on the Voice. In *Clinical Assessment of Voice*, 2nd ed.; Sataloff, R.T., Ed.; Plural Publishing, Inc.: San Diego, CA, USA, 2017; Chapter 13, pp. 221–240.
30. Boone, D.R.; McFarlane, S.C.; Berg, S.L.V.; Zraick, R.I. *The Voice and Voice Therapy*, 8th ed.; Pearson Education: Upper Saddle River, NJ, USA, 2010.
31. Linville, S.E.; Fisher, H.B. Acoustic Characteristics of Women's Voices with Advancing Age. *J. Gerontol.* **1985**, *40*, 324–330. [[CrossRef](#)]
32. Wohler, A.B.; Smith, A. Spatiotemporal Stability of Lip Movements in Older Adult Speakers. *J. Speech Lang. Hear. Res.* **1998**, *41*, 41–50. [[CrossRef](#)] [[PubMed](#)]
33. Klein, D.R. Oral soft tissue changes in geriatric patients. *Bull. N. Y. Acad. Med.* **1980**, *56*, 721–727. [[PubMed](#)]
34. Mahne, A.; El-Haddad, G.; Alavi, A.; Houseni, M.; Moonis, G.; Mong, A.; Hernandez-Pampaloni, M.; Torigian, D.A. Assessment of Age-Related Morphological and Functional Changes of Selected Structures of the Head and Neck by Computed Tomography, Magnetic Resonance Imaging, and Positron Emission Tomography. *Semin. Nucl. Med.* **2007**, *37*, 88–102. [[CrossRef](#)] [[PubMed](#)]
35. Mautner, H. A Cross-System Instrumental Voice Profile of the Aging Voice: With Considerations of Jaw Posture Effects. Ph.D. Thesis, University of Canterbury, Christchurch, New Zealand, 2011.
36. Bässler, R. Histopathology of different types of atrophy of the human tongue. *Path. Res. Pract.* **1987**, *182*, 87–97. [[CrossRef](#)]
37. Kuruvilla-Dugdale, M.; Dietrich, M.; McKinley, J.D.; Deroche, C. An exploratory model of speech intelligibility for healthy aging based on phonatory and articulatory measures. *J. Commun. Disord.* **2020**, *87*, 105995. [[CrossRef](#)]

38. Vanderwegen, J.; Guns, C.; Van Nuffelen, G.; Elen, R.; De Bodt, M. The Influence of Age, Sex, Bulb Position, Visual Feedback, and the Order of Testing on Maximum Anterior and Posterior Tongue Strength and Endurance in Healthy Belgian Adults. *Dysphagia* **2013**, *28*, 159–166. [\[CrossRef\]](#)
39. Crow, H.C.; Ship, J.A. Tongue Strength and Endurance in Different Aged Individuals. *J. Gerontol. Med. Sci.* **1996**, *51A*, M247–M250. [\[CrossRef\]](#)
40. Mortimore, I.L.; Fiddes, P.; Stephens, S.; Douglas, N.J. Tongue protrusion force and fatigability in male and female subjects. *Eur. Respir. J.* **1999**, *14*, 191–195. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Mortimore, I.L.; Bennett, S.P.; Douglas, N.J. Tongue protrusion strength and fatigability: Relationship to apnoea/hypopnoea index and age. *J. Sleep Res.* **2000**, *9*, 389–393. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Mücke, D.; Thies, T.; Mertens, J.; Hermes, A. Age-related effects of prosodic prominence in vowel articulation. In Proceedings of the 12th International Seminar on Speech Production, New Haven, CT, USA, 14–18 December 2021.
43. Chantaramanee, A.; Tohara, H.; Nakagawa, K.; Hara, K.; Nakane, A.; Yamaguchi, K.; Yoshimi, K.; Junichi, F.; Minakuchi, S. Association between echo intensity of the tongue and its thickness and function in elderly subjects. *J. Oral Rehabil.* **2019**, *46*, 634–639. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Xue, S.A.; Hao, G.J. Changes in the Human vocal tract due to aging and the acoustic correlates of speech production: A pilot study. *J. Speech Lang. Hear. Res.* **2003**, *46*, 689–701. [\[CrossRef\]](#)
45. Zharkova, N.; Hewlett, N.; Hardcastle, W.J. An ultrasound study of lingual coarticulation in /s V/ syllables produced by adults and typically developing children. *JIPA* **2012**, *42*, 193–208. [\[CrossRef\]](#)
46. Lancia, L.; Rausch, P.; Morris, J.S. Automatic quantitative analysis of ultrasound tongue contours via wavelet-based functional mixed models. *J. Acoust. Soc. Am.* **2015**, *137*, EL178–EL183. [\[CrossRef\]](#)
47. Tabain, M. Research Methods in Speech Production. In *Bloomsbury Companion to Phonetics*; Jones, M.J., Knight, R.A., Eds.; Bloomsbury: London, UK, 2013; Chapter 3, pp. 39–56.
48. Mozaffari, M.H.; Wen, S.; Wang, N.; Lee, W. Real-time automatic tongue contour tracking in ultrasound video for guided pronunciation training. In Proceedings of the VISIGRAPP 2019: 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Prague, Czech Republic, 25–27 February 2019; Volume 1, pp. 302–309. [\[CrossRef\]](#)
49. Akgul, Y.S.; Stone, C.; Maureen, K. Automatic extraction and tracking of contours. *Trans. Med. Imaging* **1999**, *18*, 1035–1045. [\[CrossRef\]](#)
50. Stone, M. A guide to analysing tongue motion from ultrasound images. *Clin. Linguist. Phon.* **2005**, *19*, 455–501. [\[CrossRef\]](#)
51. Zharkova, N.; Hewlett, N.; Hardcastle, W.J. Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control* **2011**, *15*, 118–140. [\[CrossRef\]](#)
52. Scobbie, J.M.; Lawson, E.; Cowen, S.; Cleland, J.; Wrench, A.A. *A Common Co-Ordinate System for Mid-Sagittal Articulatory Measurement*; QMU CASL Working Papers WP-20; 2011. *unpublished work*.
53. Noble, A.; Boukerroui, D.; Noble, A.; Boukerroui, D.; Noble, J.A.; Member, S.; Boukerroui, D. Ultrasound image segmentation: A survey. *IEEE Trans. Med. Imaging* **2006**, *25*, 987–1010. [\[CrossRef\]](#)
54. Strycharczuk, P.; Scobbie, J.M. Fronting of Southern British English high-back vowels in articulation and acoustics. *J. Acoust. Soc. Am.* **2017**, *142*, 322–331. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Barbier, G.; Perrier, P.; Ménard, L.; Payan, Y.; Tiede, M.; Perkell, J. Speech Planning in 4-Year-Old Children Versus Adults: Acoustic and Articulatory Analyses. In Proceedings of the Interspeech, Dresden, Germany, 6–10 September 2015.
56. Comivi Alowonou, K.; Wei, J.; Lu, W.; Liu, Z.; Honda, K.; Dang, J. Acoustic and Articulatory Study of Ewe Vowels: A Comparative Study of Male and Female. In Proceedings of the Interspeech, Graz, Austria, 15–19 September 2019; pp. 1776–1780. [\[CrossRef\]](#)
57. Ménard, L.; Toupin, C.; Baum, S.R.; Drouin, S.; Aubin, J.; Tiede, M. Acoustic and articulatory analysis of French vowels produced by congenitally blind adults and sighted adults. *J. Acoust. Soc. Am.* **2013**, *134*, 2975–2987. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Zharkova, N.; Hewlett, N.; Hardcastle, W.J. Analysing coarticulation in Scottish English children and adults: An ultrasound study. *Can. Acoust.* **2008**, *36*, 158–159.
59. Zharkova, N.; Lickley, R.; Hardcastle, W.J. Development of lingual coarticulation and articulatory constraints between childhood and adolescence: An ultrasound study. In Proceedings of the 10th International Seminar on Speech Production (10th ISSP), Cologne, Germany, 5–8 May 2014; pp. 472–475.
60. Lee, S.H.; Yu, J.F.; Hsieh, Y.H.; Lee, G.S. Relationships between formant frequencies of sustained vowels and tongue contours measured by ultrasonography. *Am. J. Speech Lang. Pathol.* **2015**, *24*, 739–749. [\[CrossRef\]](#)
61. Radisic, M. An Ultrasound and Acoustic Study of Turkish Rounded/Unrounded Vowel Pairs. Ph.D. Thesis, University of Toronto, Toronto, ON, Canada, 2014.
62. Georgetown, L.; Antolík, T.K.; Fougerson, C. Effect of Domain Initial Strengthening on Vowel Height and Backness Contrasts in French: Acoustic and Ultrasound Data. *JSLHR* **2016**, *59*, S1575–S1586. [\[CrossRef\]](#)
63. Song, J.Y. The use of ultrasound in the study of articulatory properties of vowels in clear speech. *Clin. Linguist. Phon.* **2017**, *31*, 351–374. [\[CrossRef\]](#)
64. Baghban, K.; Zarifian, T.; Adibi, A.; Shati, M.; Derakhshandeh, F. The quantitative ultrasound study of tongue shape and movement in normal Persian speaking children. *Int. J. Pediatr. Otorhinolaryngol.* **2020**, *134*, 110051. [\[CrossRef\]](#)

65. Zharkova, N. A normative-speaker validation study of two indices developed to quantify tongue dorsum activity from midsagittal tongue shapes. *Clin. Linguist. Phon.* **2013**, *27*, 484–496. [CrossRef]
66. Articulate Instruments Ltd. *Ultrasound Stabilisation Headset Users Manual*; Articulate Instruments Ltd.: Musselburgh, UK, 2008.
67. Articulate Assistant Ltd. *Articulate Assistant Advanced Ultrasound Module User Manual*; Articulate Instruments Ltd.: Musselburgh, UK, 2014.
68. Articulate Instruments Ltd. *SyncBrightUp Users Manual*; Articulate Instruments Ltd.: Musselburgh, UK, 2010.
69. Wrench, A.A.; Scobbie, J.M. High-speed Cineloop Ultrasound vs. Video Ultrasound Tongue Imaging: Comparison of Front and Back Lingual Gesture Location and Relative Timing. In Proceedings of the 8th International Seminar on Speech Production (ISSP), Alsace, France, 8–12 December 2008.
70. Dokovova, M.; Sabev, M.; Scobbie, J.M.; Lickley, R.; Cowen, S. Bulgarian vowel reduction in unstressed position: An ultrasound and acoustic investigation. In Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia, 5–9 August 2019; pp. 2720–2724.
71. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect. Notes Comput. Sci.* **2015**, *9351*, 234–241.
72. Kirkham, S.; Nance, C. An acoustic-articulatory study of bilingual vowel production: Advanced tongue root vowels in Twi and tense/lax vowels in Ghanaian English. *J. Phon.* **2017**, *62*, 65–81. [CrossRef]
73. Kisler, T.; Reichel, U.; Schiel, F. Multilingual processing of speech via web services. *Comput. Speech Lang.* **2017**, *45*, 326–347. [CrossRef]
74. Boersma, P.; Weenink, D. Praat: Doing Phonetics by Computer. 2012. Available online: <https://www.fon.hum.uva.nl/praat/> (accessed on 28 December 2021).
75. Mozaffari, M.H.; Lee, W.S. Domain adaptation for ultrasound tongue contour extraction using transfer learning: A deep learning approach. *J. Acoust. Soc. Am.* **2019**, *146*, EL431–EL437. [PubMed]
76. Zhu, J.; Styler, W.; Calloway, I. A CNN-based tool for automatic tongue contour tracking in ultrasound images. *arXiv* **2019**, arXiv:1907.10210.
77. Watson, P.J.; Munson, B. A comparison of vowel acoustics between older and younger adults. In Proceedings of the ICPhS XVI, Saarbrücken, Germany, 6–10 August 2007; pp. 561–564.
78. Torre, P., III; Barlow, J.A. Age-related changes in acoustic characteristics of adult speech. *J. Commun. Disord.* **2009**, *42*, 324–333. [CrossRef]
79. Rastatter, M.P.; Jacques, R.D. Formant frequency structure of the aging male and female vocal tract. *Folia Phoniatr.* **1990**, *42*, 312–319. [CrossRef]
80. Sonies, B.C.; Stone, M.; Shawker, T. Speech and Swallowing in the Elderly. *Gerodontology* **1984**, *3*, 115–123. [CrossRef]
81. Lee, W.S. Articulatory-Acoustical Relationship in Cantonese Vowels. *Lang. Linguist.* **2016**, *17*, 477–500. [CrossRef]
82. Mielke, J. An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *J. Acoust. Soc. Am.* **2015**, *137*, 2858–2869. [CrossRef]
83. Turgeon, C.; Trudeau-Fisette, P.; Fitzpatrick, E.; Ménard, L. Vowel intelligibility in children with cochlear implants: An acoustic and articulatory study. *Int. J. Pediatr. Otorhinolaryngol.* **2017**, *101*, 87–96. [CrossRef]