The VoxSenes project: a study of segmental changes and rhythm variations on European Portuguese aging voice

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

[1]Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Aveiro, Portugal
[2]School of Health Sciences, University of Aveiro, Portugal
[3]Dep. Electronics, Telecommunications and Informatics, University of Aveiro, Portugal
[4]Center for Health Technology and Services Research, University of Aveiro, Portugal
[5]Department of Education and Psychology, University of Aveiro, Aveiro, Portugal
[6]Institute of Biomedicine, University of Aveiro, Portugal

coliveira@ua.pt, rita.valente@ua.pt, lucianapereira@ua.pt, fabiodaniel@ua.pt, pmartins@ua.pt, sss@ua.pt, ajst@ua.pt

Abstract

The process of aging is generally associated with a number of changes in physiological, cognitive, psychological and social domains, including modifications on vocal quality of individuals. This paper presents a recent project - VoxSenes - that intends to bridge knowledge gaps in the speech changes due to aging. A deeper knowledge on how speech changes with age is essential for the development of automatic speech recognition systems suitable for older's voices and for clinical assessment of speech disorders. The VoxSenes project aims to study aging voice at segmental and suprasegmental level. The variation of acoustic parameters were analysed through audio speech samples and articulatory parameters were investigated over ultrasound tongue imaging. The most relevant age-related results of this project include: an increase in vowel duration, an approximation of F0 between genders, a centralization of the acoustic vowel space for males and an increase in speech pauses. The unsupervised method already developed to extract tongue contours from ultrasound tongue images provides the required data for an automatic analysis of relevant parameters to assess speech changes on vowel production.

An analysis of articulatory space of European Portuguese vowels is ongoing for speakers of different age groups, as well as a longitudinal study of age-related changes in speech rhythm.

Index Terms: Aging speech, Acoustic, European Portuguese, Speech production

1. Introduction

Population aging is one of the greatest triumphs of modern society, but also one of its major challenges [1]. More than in other developed countries, the Portuguese population has been aging (between 1970 and 2018, the percentage of people aged 65 and over increased from 9.7% to 21.8% [2, 3], both because of increased life expectancy and the decrease in fertility rates [1].

The process of aging is generally associated with physiological, cognitive, psychological and social changes, including modifications on individuals' vocal parameters. From childhood to old age, the speech production mechanism undergoes numerous anatomical and physiological changes in the respiratory, laryngeal and supralaryngeal system [4, 5, 6, 7]. The differences between men and women regarding the timing and extent of age-related changes are substantial [5, 6, 7]. All of these physical changes result in complex variations in the acoustic properties of speech (F0, stability of vocal fold vibration, spectral noise, speech tempo and formant frequencies), which are influenced by several variables (e.g., speaker-related factors or speech-material-related factors) [6].

A general finding is that older adults demonstrate a higher variation of acoustic features than younger adults. Most research has reported that F0 increases in males and decreases in females with aging [8]. Tremor and increased hoarseness, that have been associated with the aged voice, seems to be the result of a decline of F0 or amplitude stability [9]. It has also been noted that the age of the speaker strongly affects speaking and reading rates. Only relatively few studies have investigated age-related changes in the vocal tract resonance features. Most of them referred a general lowering of formants with age [10, 11] attributed to a lengthening of the vocal tract, caused by a lowering of the larynx, of the tracheobronchial tree and of the lungs, and by a growth of the facial skeleton. There also seems to be a trend towards vowel centralization [12, 8], but in some cases these changes have only been observed for particular vowels and speakers [12]. The longitudinal studies' findings (e.g., [13]) are in line with results of studies based on between-group comparisons (both F0 and F1 decrease with increasing age in the same speaker).

Additionally, changes in speech rhythm of old speakers have been scarcely examined [14].

For European Portuguese (EP) there is almost no data on the acoustic correlates of aging. Pellegrini et al. [14] conducted the first attempt to identify the most salient differences between older and younger adult speech in terms of acoustic features for EP, mostly to understand their impact on speech recognition performance. In a pilot study, our team [15] addressed effects of age and gender on formant frequencies of oral vowels produced by EP old speakers. We also investigated acoustic characteristics of EP children's speech [16]. So far the few studies on EP vowels acoustics [17, 18] have only provided information about vowels produced by young adult speakers.

Even though acoustic methods are a valuable tool in the study of speech sounds, the integration of acoustic with articulatory data could facilitate a comprehensive account of

10.21437/IberSPEECH.2021-29

anatomic–acoustic relationships. However, the existing studies mainly focused on the developmental period from infancy to adulthood [19]. Many instrumental techniques have been used in this field (e.g., Magnetic Resonance Imaging - MRI and ultrasound - US). US imaging is currently gaining popularity as a research tool in speech production studies because it is safe, non-invasive, simple and cost-effective. Moreover, it allows a simple synchronized acquisition of data (e.g., with audio). Although, acquisition and analysis of articulatory data is extremely demanding and poses several challenges.

The aim of the VoxSenes project is to study the speech changes due to aging, both at the segmental level as well as at suprasegmental level, by analyzing the variation of acoustic and articulatory parameters. A deeper knowledge on how speech changes with age is essential for the development of automatic speech recognition systems suitable for older voices (personalized reading aids and voice prostheses) and to clinical assessment of speech disorders.

2. Methodology

To achieve the defined aims, the VoxSenes project was organized in three different phases. In the first phase of this project, following up on pilot studies carried out by our team, the acoustic EP vowels produced by a large number of speakers, divided into different age groups, were analyzed. The relationship between the production of vowels (tongue shape and motion), the changes on formant frequencies and age were addressed in an ultrasound study to be held in the second task of the project. The articulatory basis of previously acoustic findings concerning speech aging were studied through the use of US tongue imaging synchronized with audio. The age-related changes of rhythm were approached in a third phase of the project, with the purpose of evaluating rhythmic changes throughout a certain span of time for a same individual (longitudinal study). The methodological procedures of phase 1 and 2 involve the recruitment of participants of different age groups. The procedures were submitted to Ethical Approval. Participants were fully informed about the goals, procedures and risks involved in research and had to give their consent to participate (Informed Consent). The researchers also guarantee the confidentiality and anonymity of the data, which means that identifying information will not be made available to anyone who is not directly involved in the study. The inclusion criteria were: age over 18 years old; European Portuguese as mother tongue. The exclusion criteria were: previous history of speech-language impairments, head/ neck cancer and/or neurological disorders; upper respiratory tract infection; currently smokers or had smoked within the previous 5 years; poor general health; and use of hearing aids. The participants were recruited through personal contacts and/or snowball technique in the community, and in Senior Universities from the center of Portugal.

2.1. VoxSenes Phase 1

The speech sample consisted of 36 real words, with the vowels of the EP [i], [e], [ɛ], [a], [o], [ɔ], [u] in stressed position and the vowels [ɐ] and [ɨ] in unstressed position. For each vowel, four different words were selected. Each vowel was produced in a disyllabic sequence, mostly CV.CV (C-consonant, V-vowel) where C was a voiced/ voiceless stop consonant or a voiced/ voiceless fricative consonant. The stimuli were embedded in a carrier sentence. The participants were also instructed to describe the "Cookie Theft" picture [20, 21] from the Boston Diagnostic Aphasia Examination at comfortable pitch and loudness level. All the recording took place in quiet rooms, using an AKG C535 EB cardioid condenser microphone connected to an external 16-bit sound system (PreSonus Audio-BoxTM USB), at a sampling rate of 44100 Hz. The sentences were randomized and presented individually using the software SpeechRecorder [22] with pictures and the orthographic word simultaneously. The recruited participants read the sentences 3 times at comfortable pitch and loudness level, after familiarizing with the sentences.

The recorded data from vowel production were automatically segmented at phoneme level using WebMAUS General for Portuguese language (PT) [23] and then imported into Praat [24]. Four trained analyzers manually checked the accuracy of the vowel boundaries. The acoustic parameters of the vowels - F0, median F0, F1, F2 and duration - were automatically extracted from the data set using Praat scripts. The recorded data from the picture description task were segmented for pauses over length 250 ms using a Praat scipt [25]. The Praat scripts ProsodyDescriptor [26] and the BeatExtractor [27] were used to extract the acoustic measures: total speech duration, percent pause time, mean pause duration, speaking rate, articulatory rate, speaking F0 and harmonic-to-noise ratio (HNR).

The research team is now determining the potential usefulness of the dynamic acoustic proprieties of EP vowels as carriers of age classification, a work that followed the one already carried out on the age effects in static cues (formant frequencies and F0) [15, 28, 29].

2.2. VoxSenes Phase 2

The synchronous acquisition of US images and speech sounds using Articulate Assistant Advanced software (AAA) [30] took place in a quiet room, using an endocavitary probe (65EC10EA) with 90º field of view positioned under the participants' chin using a stabilization helmet [31]. US was collected using a Mindray DP6900 at a frame rate of 60 Hz. Audio was collected with a Philips SBC ME400 microphone connected to an external sound system (UA-25 EX USB). The corpus consisted of 9 repetitions for each of the 9 EP oral vowels ([i], [e], [ɛ], [a], [o], [ɔ], [u], [ɐ] and [ɨ]). Corpus acquisition began with the production of the sequence /tatatata/ to assess sound and image synchronization and with swallowing saliva for hard palate delineation. The recorded data was collect as video and audio synchronized with SyncBrightUp unit [32], and the audio was automatically segmented, at phoneme level, using WebMAUS General [23].

Based on the acoustic midpoint of the vowels, the corresponding images were selected and processed using an unsupervised method to extract points-of-interest in the tongue. The method uses a radial sweep approach [33] and collects all the pixel intensities. The highest intensity point is extracted and the highest y coordinate is considered to represent the highest point of the tongue. The x-coordinate reflects the front-back position of the tongue in the x-y coordinate system. To allow intra- and inter-subject comparisons, normalization procedures were implemented. Concerning the analysis, the researchers intend to implement qualitative and quantitative analysis. The tongue surface tracing will be submitted to smoothing spline ANOVA (SSANOVA) [34] to find the best-fit curve across repetitions, allowing comparisons of tongue contours.

2.3. VoxSenes Phase 3

Longitudinal data of EP speakers (e.g., politicians or journalists) was obtained. The speech data were selected from national television archives. The selection of speech material for analysis depend mainly on the following criteria: similar type (e.g., reading or interviews) and duration of speech samples; quality of recordings and homogeneity of recording type across speakers and samples; voice quality of the speaker for enabling acoustic analysis.

For all sample, the vowel onsets were marked semi-automatically using the Praat script BearExtractor [27]. Fourteen rhythm and intonation parameters were extracted using the Praat script ProsodyDescriptor [26].

3. Results

3.1. VoxSenes Phase1

The results of Phase 1 were published, so far, in 1 journal paper ([35]) and 2 conference papers ([28]; [36]). The dynamic vowel studies result in one poster [29] and one position paper [37].

113 native Portuguese speakers (56 men and 57 women) from the central region of Portugal, aged between 35 and 97, participated in this phase of the VoxSenes project. The participants were divided into 4 age groups: [35-49] (15 men, 15 women), [50-64] (15 men, 15 women), [65-79] (15 men, 16 women), and ≥ 80 (11 men, 11 women).

The acoustic data revealed that the duration of all vowels increased with aging, and that the older speakers presented the longest vowel duration. F0 decreased in male until the age group [50-64] and started to increase after that age, with a more pronounced increase in the group ≥ 80, which presented the highest mean value of F0. For female speakers, the opposite tendency was observed, with an F0 increase until the age group [50-64] and a sharp decrease after this age. The age group ≥ 80 presented the highest mean value of F0. We also observed a general lowering of F1 and F2 frequencies for women in all stressed vowels; for males, changes observed in F1 and F2 were consistent with vowel centralization. There was no evidence of F3 decrease with age.

Regarding rhythm and intonation study, the most consistent age-related effects were an increase in speech pauses (both duration and percent time), mainly in men, and a HNR decrease in women. The articulatory rate differences between male and female decreased with age. Speaking F0 presented a similar tendency observed in the acoustic vowel data for male, with a decrease until the age group [65-79] and a rise after that age.

Concerning vowel dynamics, the results already achieved showed that dynamic measurements of F1-F3 result in better classification performance of senior/non-senior. Duration was also reconfirmed as an important predictor of age, just like in other studies of our team with static cues [35].

3.2. VoxSenes Phase2

The results already achieved on the articulatory changes in lifetime concern the accuracy of the method to extract tongue contours. The results were published in [38] and [39].

Data revealed that, in general, the method developed presented high accuracy mainly for the central regions of the tongue, where the highest point of the tongue is typically located. This method is currently being used, in an ongoing study with a larger sample, to characterize the articulatory changes of vowels with aging.

3.3. VoxSenes Phase3

Concerning phase 3 of the VoxSenes project, the researchers are currently selecting the speech data samples, more specifically interviews of two males in different ages, to allow the analysis of age effect in rhythm and intonation parameters.

4. Impact

Vocal aging is a multidimensional concept, making the study of older speech complex and challenging. The VoxSenes project intends to provide an opportunity for trying out a number of acoustic and articulatory research methods, in order to contribute to the increase of the phonetic knowledge concerning the properties of speech and its changes with age.

The results already achieved intend to have an impact on a better understanding of cross-linguistic similarities and in language-specific features of vowel aging. Furthermore, the new databases created in this project allow the characterization of vowel production in the normative aging process, being a reference for clinical assessment and intervention of different speech disorders. Concerning speech technology, data collected on the current project provides information that can have a positive impact on better age recognizers or classifiers, as well as in more natural-sounding synthesis of speaker age, which could be useful to improve the quality of life of older people [40, 41].

5. Acknowledgements

This research project is financially supported by POCI-01-0145-FEDER-03082 (funded by FEDER, through COMPETE2020 - Programa Operacional Competitividade e Internacionalização (POCI), and by national funds (OE), through FCT/MCTES). The researchers also intend to acknowledge the grant SFRH/BD/115381/2016 and IEETA (UIDB/00127/2020). We are very grateful to the institutions for having made possible the data collection, and also to all adults who contributed as speakers.

6. References

[1] W. He, D. Goodkind, and P. R. Kowal, "An aging world: 2015," *International Population Reports*, vol. P95/16-1, 2016.

[2] Statistics Portugal, "Estimativas de População Residente em Portugal - 2018 (Estimates of resident population in Portugal - 2018)," *Destaque: informação à comunicação social*, 2019.

[3] ——, "Envelhecimento da população residente em Portugal e na União Europeia (Aging of the resident population in Portugal and the European Union)," *Destaque: informação à comunicação social*, 2015.

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

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

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

[7] K. Makiyama and S. Hirano, "Aging Voice," Singapore, 2017.

[8] P. Torre III and J. A. Barlow, "Age-related changes in acoustic characteristics of adult speech," *Journal of Communication Disorders*, vol. 42, pp. 324–333, 2009.

[9] S. E. Linville, "The Sound of Senescence," *Journal of Voice*, vol. 10, no. 2, pp. 190–200, 1996.

[10] P. J. Watson and B. Munson, "A comparison of vowel acoustics between older and younger adults," in *ICPhS XVI*, Saarbrücken, 2007, pp. 561–564.

[11] S. A. Xue and G. J. Hao, "Changes in the Human vocal tact due to aging and the acoustic correlates of speech production: a pilot study," *J Speech Lang Hear Res*, vol. 46, no. 3, pp. 689–701, 2003.

[12] 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.

[13] J. Harrington, S. Palethorpe, and C. I. Watson, "Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers," in *INTERSPEECH*, Belgium, 2007, pp. 2753–2756.

[14] 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.

[15] 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.

[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, Spain, 2012, pp. 129–138.

[17] M. R. D. Martins, "Análise acústica das vogais orais tónicas em Português," *Boletim de Filologia*, vol. 22, pp. 303–314, 1973.

[18] 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.

[19] H. K. Vorperian and R. D. Kent, "Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data," *J Speech Lang Hear Res*, vol. 50, no. 6, pp. 1510–1545, 2007. [Online]. Available: http://jslhr.asha.org/cgi/content/abstract/50/6/1510

[20] H. Goodglass and E. Kaplan, *The Assessment of Aphasia and Related Disorders*, 2nd ed. Philadelphia, PA.: Lea and Febiger, 1983.

[21] E. E. Morgan and M. Rastatter, "Variability of voice fundamental frequency in elderly female speakers." *Perceptual and motor skills*, vol. 63, no. 1, pp. 215–218, 1986.

[22] C. Draxler and K. Jänsch, "SpeechRecorder (3.12.0)," 2017.

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

[24] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," University of Amsterdam, 2012. [Online]. Available: http://www.praat.org/

[25] N. H. de Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior Research Methods*, vol. 41, no. 2, pp. 385–390, may 2009.

[26] P. A. Barbosa, "Semi-automatic and automatic tools for generating prosodic descriptors for prosody research," in *TRASP*, vol. 13, no. 2, Aix-en-Provence, 2013, pp. 86–89.

[27] ——, "Automatic duration-related salience detection in Brazilian Portuguese read and spontaneous speech," in *Speech Prosody*, Chicago, 2010.

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

[29] L. Albuquerque, A. Teixeira, C. Oliveira, and D. Figueiredo, "The effect of dynamic acoustic cues on age classification," in *SPPL2020: 2nd Workshop on Speech Perception and Production across the Lifespan (Poster)*, 2020, p. 81.

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

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

[32] ——, "SyncBrightUp users manual," Edinburgh, UK, 2010.

[33] 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.

[34] L. Davidson, "Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance," *J. Acoust. Soc. Am.*, vol. 120, no. 1, pp. 407–415, 2006.

[35] 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.

[36] L. Albuquerque, A. R. S. Valente, A. Teixeira, C. Oliveira, and D. Figueiredo, "Acoustic changes in spontaneous speech with age," in *VIII Congreso Internacional de Fonética Experimental*, Girona, 2021.

[37] L. Albuquerque, C. Oliveira, A. Teixeira, and D. Figueiredo, "Eppur si muove: Formant dynamics is relevant for the study of Speech Aging Effects," in *14th BIOSIGNALS*, Online, 2021, p. in press.

[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 (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 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*, Online, 2020, p. Poster.

[40] S. O. Sadjadi, S. Ganapathy, and J. W. Pelecanos, "Speaker age estimation on conversational telephone speech using senone posterior based i-vectors," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 5040–5044. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7472637/

[41] M. Yue, L. Chen, J. Zhang, and H. Liu, "Speaker age recognition based on isolated words by using SVM," in *CCIS2014*, 2014, pp. 282–286.