Baltic J. Modern Computing, Vol. 10 (2022), No. 3, 360-371
https://doi.org/10.22364/bjmc.2022.10.3.09

Estonian Elderly Speech Corpus – Design, Collection and Preliminary Acoustic Analysis

Einar MEISTER, Lya MEISTER

Tallinn University of Technology, Department of Software Science,
Akadeemia tee 15A, 12618, Estonia

einar.meister@ttu.ee, lya.meister@ttu.ee

Abstract. Elderly speech has been found challenging for automatic speech recognition systems due to the lack of suitable training data and due to age-related physiological changes affecting the acoustic characteristics of voice. The paper introduces the collection of the speech corpus of Estonian elderly speakers aged over 60 years and reports preliminary results of acoustic analysis of some prosodic characteristics. The corpus contains the recordings of spontaneous speech by 100 men and 100 women, on average 29 minutes of speech from each speaker with transcriptions. Acoustic analysis of elderly speech revealed that with age (1) F0 increases significantly in males and slightly decreases in females, (2) speech and articulation rates decrease in females, but not in males, (3) utterances are getting shorter, and (4) pauses are getting longer in females and shorter in males.

Keywords: speech corpus, elderly speech, acoustic analysis, fundamental frequency, speaking rate

1. Introduction

Since the mid-1990s, the collection of different speech corpora has been a systematic activity at the Language Technology Lab at Tallinn University of Technology. The corpora are necessary mainly for speech technology development, especially for the training the acoustic models of the ASR systems, and some of them are used for experimental-phonetic studies. By 2022, the amount of speech data available for the training of the acoustic models of Estonian ASR is ca 800 hours of manually transcribed speech. Regular growth of the amount of training data and the implementation of DNN-based methods in acoustic and language modelling have significantly contributed to the reduction of Estonian ASR errors. For example, the word error rate (WER) in the case of radio news has dropped from 28.5% in 2010 to 8.5% in 2022. However, Estonian ASR does not perform so well in the case of spontaneous speech (WER = 17.6%) and elderly speech (WER = 18.8%). In addition to the differences between read and spontaneous speech, these results can be explained by the fact that speakers aged over 60 years are rarely present in the training corpora, and by the age-related changes of voice and speech characteristics that diverge from the characteristics of middle-aged speakers. As a result, the acoustic models trained with speech samples from middle-aged speakers are not suitable to recognize speech produced by elderly voices (Hämäläinen et al., 2014). A comprehensive review of 17 studies concluded that ASR systems have a higher WER in recognizing the speech of older (60–80 years old) adults compared to WER of younger (20–60 years old) adults, and it increases with age (Werner et al. 2019). The review also confirmed that the older adult WER values can be lowered if the training corpus employs samples of older adults’ speech.

To improve elderly speech recognition, several speech corpora have been collected. JASMIN-CGN (Cucchiarini et al., 2006) corpus includes 25 hours of Dutch and Flemish elderly speech, EASR corpora include samples of European Portuguese, French, Hungarian, and Polish elderly speech (76 to 205 hours of speech from 328 to 986 speakers, depending on the language) (Hämäläinen et al., 2014), Bengali corpus of elderly speech comprises samples of 60 speakers (Das et al., 2012), S-JNAS corpus (Baba et al., 2002) and a more recent corpus (Fukuda et al., 2020) include 133 hours (301 speakers) and 22 hours (221 speakers) of Japanese elderly speech, respectively.

Aging involves several degenerative changes in the human body including the parts of the speech production mechanism – the respiratory system, larynx, and the oral cavity (Linville, 2001). The undergoing physiological changes (e.g., loss of elasticity in the respiratory system, calcification of the laryngeal tissues, stiffening of the vocal cords, degeneration of intrinsic muscles) affect speech production in several ways: lung pressure decreases, the instability of the vocal fold vibrations increases, loss of tongue strength, changes in dimensions of the oral cavity. Differences in several acoustic parameters between elderly and middle-aged voices have been reported for different languages, including changes in formants, fundamental frequencies (rising in males and lowering in females), voice quality changes (increased breathiness, jitter, and shimmer), and slower speaking rate (e.g., Albuquerque et al., 2019, Torre and Barlow, 2009, Vipperla et al., 2010, Eichhorn et al., 2017, Xue and Hao, 2003). Vocal ageing, in addition to the physiological changes, is influenced also by several other factors, such as lifestyle, smoking, alcoholism, and food habits (Gorham-Rowan and Laures-Gore, 2006; Linville, 2001, Vipperla et al., 2010).

The paper introduces the design and collection of the speech corpus of Estonian elderly speakers aged over 60 years. It also reports the preliminary results of acoustic analysis of some prosodic characteristics – fundamental frequency (F0) and speech tempo – depending on age and gender.

2. Elderly speech data collection

2.1. Corpus design

The elderly speech corpus targets to extend the existing Estonian speech corpora with speech samples from the age group almost not presented in the current corpora (e.g., Meister et al., 2003; Meister et al., 2012; Meister and Meister, 2014). The corpus will be used for training speech recognition systems and socio-phonetic studies addressing the changes in voice and speech characteristics of this age group.

The number of speakers in the corpus is 200 balanced by gender and age groups. The following requirements for speaker selection were set:

- age over 60 years,
- no obvious voice pathology or disability,
- able to hear and speak normally,
- native or near-native speakers of Estonian.

2.2. Corpus content

Unlike several elderly speech corpora collected in other languages (Hämäläinen et al., 2012; Hämäläinen et al., 2014; Fukuda et al., 2020) that contain samples of read speech, our corpus contains spontaneous speech samples only. Speech was elicited during the interviews/conversations guided by one of the authors (most interviews were performed by the second author). The interviewees were encouraged to choose their preferred topics for storytelling, however, most speakers expected to be guided by the hints or questions from the interviewer. The topics cover a wide range, most frequently addressing memories of childhood, school time, working life, family, hobbies, traveling, etc. The interviewers avoided sensitive personal questions such as health problems, economic survival or politics, and religion unless these topics were brought up by the speakers themselves. E.g., COVID-19 and personal experience with it were addressed by several subjects.

2.3. Speaker recruitment

Calls for participation were distributed in the university, in several nursing homes and day-care centres for the elderly and within authors' personal and professional networks. Majority of the subjects were recruited in the capital area, some from South-Estonia (Tartu and Võru) and some from Western Estonia (Pärnu). All speakers participated voluntarily and were not awarded for their contribution. Before the recording session the subjects signed a consent form; in addition, from each subject the following data was collected: age, gender, education, mother tongue, other languages studied, place of living in early childhood, and current place of living. Ideal gender balance within the age groups, however, was rather difficult to achieve (see Table 1). In all age groups, the majority of speakers are with higher education (see Table 2) as they happened to be more vital and cooperative and realized more easily the purpose of data collection. In general, it was easier to recruit female speakers than males.

Table 1. Distribution of speakers by age and gender

Age group	Male	Female	Subtotal
60–69	37	33	70
70–79	39	31	70
≥ 80	24	36	60
Total	100	100	200

Table 2. Distribution of speakers by education and gender

Education level	Male	Female	Subtotal
Basic/vocational	7	9	16
Secondary/vocational	30	43	73
Higher	63	48	111

2.4. Recording procedure

In most cases, recording sessions took place at subjects' locations (in a quiet room at a nursing home or a day-care centre or at a working place or home) and in some cases in our recording studio. The recordings were carried out using a mobile recording set including a portable digital recorder (M-Audio Microtrack 24/96 or Sound Devices MixPre-3) and two cardioid lavalier microphones (Electro-Voice). Both the interviewee and the interviewer wore the microphone ca 20 cm from the lips. The recordings were made by the interviewer, no other persons were present during the recording session. The signals were stored in WAV format (two channels, sampling at 44.1 kHz, 16 bit) and a backup copy was stored into a laptop computer immediately after the recording session.

The average duration of a recording session was around 40 minutes (including explanations and setting up the recording system) whereas the average recording time was ca 28.5 minutes. However, the actual duration of the recordings varies from 9 to 65 minutes in males and from 16 to 82 minutes in females (Fig. 1.). In total, the corpus includes ca 95 hours of speech.

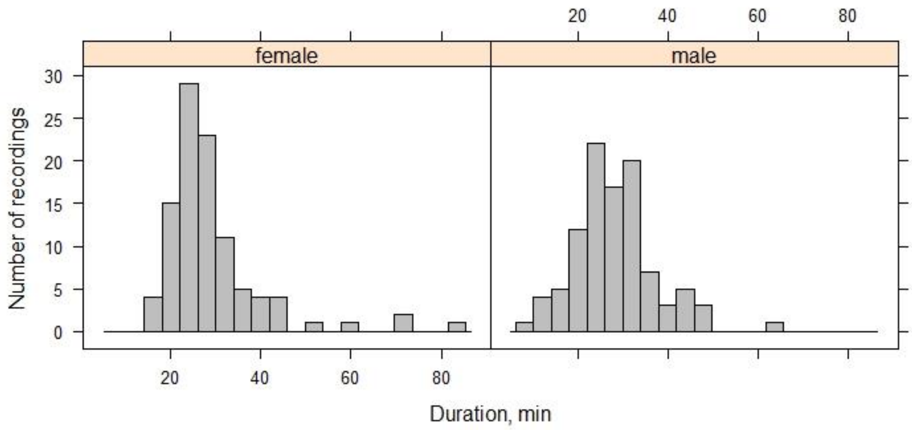

Fig. 1. Male and female histograms of recording duration.

2.5. Transcripts

All recordings were first transcribed using the lab's web service http://bark.phon.ioc.ee/webtrans/ implementing the Estonian ASR system (Alumäe et al., 2018) based on the Kaldi toolkit (Povey et al., 2011). The web service produces an automatic transcription of the uploaded speech recordings, in addition, it also performs automatic punctuation restoration and speaker diarization. Transcriptions are available in several formats including one for Transcriber http://trans.sourceforge.net/. WER of the system varies from 8.1% to 22.7% depending on the quality of the recording and speech style.

In the next step, for each recording up to ca 20 minutes of the automatic transcript was manually checked and corrected in Transcriber. Consequently, the total amount of the transcribed corpus is ca 70 hours.

3. Acoustic analysis of prosodic features

3.1. Materials

A subset of the corpus – 3-5 minutes of monologue speech (ca 100 utterances per speaker) from 100 females and 74 males (at the time of writing the manuscript, the manual corrections of the automatic transcriptions of 26 male speakers were not yet completed) – was extracted from the full corpus. For the acoustic analyses, the signals were further processed using the web service for automatic segmentation https://bark.phon.ioc.ee/autosegment2/ (Alumäe et al., 2018) which creates time-aligned word- and phone-level segments stored in Praat's (Boersma and Weenink, 2022) TextGrid-files. Next, the syllable boundaries and labels were added to TextGrids with a customized Praat script (Lippus, 2015). The subjects were grouped into three age groups: (1) 60–69 years, (2) 70–79 years, and (3) 80+ years.

3.2. Methods

The prosodic characteristics – fundamental frequency (F0), utterance and pause durations, and speech and articulation rates – were measured using a custom Praat script.

Speaking rate reveals the amount of speech produced per unit of time and is characterised by two measures: (1) speech rate – calculated as the number of syllables divided by utterance duration including disfluencies, and (2) articulation rate – calculated as the number of syllables divided by utterance duration without disfluencies.

Linear mixed effect models for F0 and the temporal characteristics were fitted using the *lme4* packages (Bates et al., 2020) in *RStudio* (RStudio Team, 2020). In the models, Age group, Gender, and Age group * Gender interaction were included as the fixed effects, the random effects included Subject and Utterance intercepts and slopes for Age group.

3.3. Results

3.3.1. Fundamental frequency (F0)

Fig. 2 plots the measured F0 mean values, Table 3 and 4 provide the model output and the model-predicted F0 means with confidence intervals, respectively.

Table 3. The fixed-effects output of the *lmer*-model for F0 (significant effects are in bold).

Predictor	Estimate	Std. error	df	t	p
(Intercept)	**189.606**	**3.100**	**69.843**	**61.163**	**<0.001**
Age group 2	-2.076	5.342	99.802	-0.389	0.698
Age group 3	-3.574	5.179	79.881	-0.690	0.492
Gender male	**-70.380**	**4.452**	**69.726**	**-15.808**	**<0.001**
Age group 2 * Gender male	7.763	7.636	100.212	1.017	0.312
Age group 3 * Gender male	16.689	9.082	63.226	1.838	0.071

Table 4. Model-predicted F0 means and 90% confidence intervals in Hz in three age groups.

Gender	Age group 1		Age group 2		Age group 3	
	mean	90% CI	mean	90% CI	mean	90% CI
Female	189.61	184.5–194.7	187.53	180.4–194.7	186.03	179.2–192.9
Male	119.23	114.0–124.5	124.91	117.6–132.2	132.34	121.3–143.4

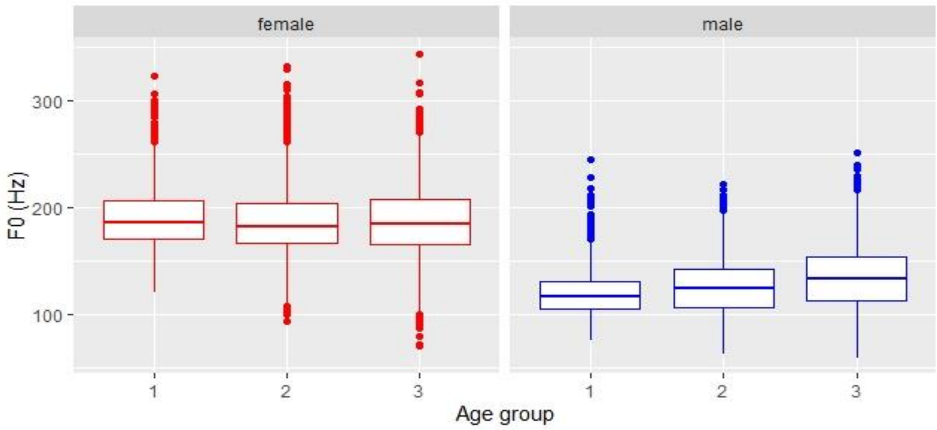

Fig. 2. Boxplots of the measured F0 means for female and male speakers in three age groups.

Analysis of variance (ANOVA) and TukeyHSD post-hoc tests of F0 means revealed significant differences across the age groups ($F = 186.2$, $p < 0.001$) in male data, for all pair-wise comparisons $p < 0.001$. In females, age group differences are significant ($F = 11.04$, $p < 0.001$) between the age groups 1 and 2, and 1 and 3 (for both $p < 0.001$), but not between the age groups 2 and 3 ($p = 0.98$).

3.3.2. Duration of utterances and pauses

The measured utterance and pause durations are visualized in Fig. 3 and 4, respectively. A linear mixed effects analysis of utterance and pause durations showed significant main effects of age group, gender and all interactions (see Tables 5 and 6); Table 7 shows the predicted utterance and pause durations and 90% confidence intervals.

Table 5. The fixed-effects output of the *lmer*-model for utterance duration.

Predictor	Estimate	Std. error	df	t	p
(Intercept)	**2.62**	**0.077**	**61.51**	**34.211**	**< 0.001**
Age group 2	-0.114	0.102	108.56	-1.112	0.268
Age group 3	**-0.483**	**0.111**	**85.32**	**-4.339**	**< 0.001**
Gender male	**-0.271**	**0.057**	**2934.6**	**-4.674**	**< 0.001**
Age group 2 * Gender male	**-0.297**	**0.082**	**2960.9**	**-3.626**	**< 0.001**
Age group 3 * Gender male	**0.287**	**0.098**	**2252.7**	**2.912**	**< 0.01**

Table 6. The fixed-effects output of the *lmer*-model for pause duration.

Predictor	Estimate	Std. error	df	t	p
(Intercept)	**0.772**	**0.027**	**61.06**	**29.13**	**< 0.001**
Age group 2	-0.053	0.04	94.55	-1.346	0.181
Age group 3	**0.096**	**0.043**	**84.75**	**2.244**	**0.027**
Gender male	**0.04**	**0.02**	**2897**	**1.975**	**0.048**
Age group 2 * Gender male	**0.081**	**0.029**	**2768.3**	**2.824**	**0.005**
Age group 3 * Gender male	**-0.151**	**0.035**	**4057.2**	**-4.348**	**< 0.001**

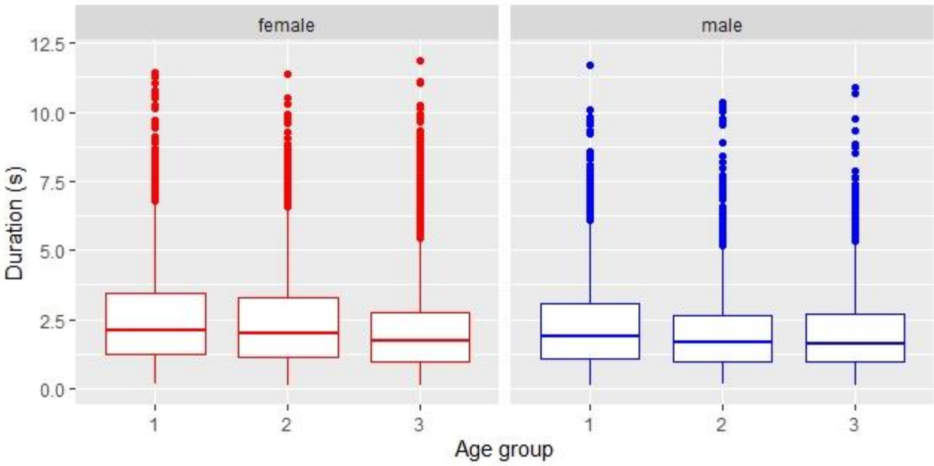

Fig. 3. Utterance durations for female and male speakers in three age groups.

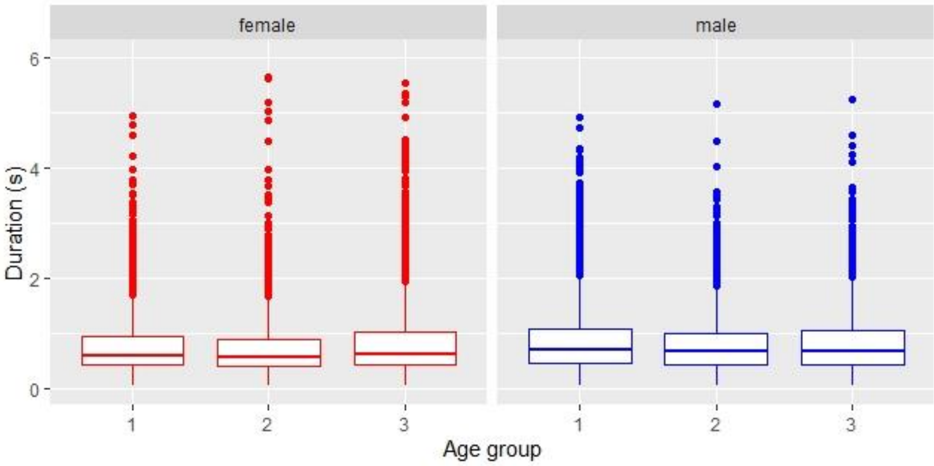

Fig. 4. Pause durations for female and male speakers in three age groups.

Table 7. Model-predicted utterance and pause durations and 90% confidence intervals (in seconds) in three age groups.

	Gender	Age group 1		Age group 2		Age group 3	
		mean	90% CI	mean	90% CI	mean	90% CI
Utterance	Female	2.62	2.50– 2.75	2.51	2.40–2.62	2.14	2.01–2.27
	Male	2.35	2.22–2.48	1.94	1.83–2.05	2.15	1.99–2.32
Pause	Female	0.77	0.73–0.82	0.72	0.67–0.77	0.87	0.81–0.92
	Male	0.81	0.77–0.86	0.84	0.79–0.89	0.76	0.69–0.82

3.3.3. Speaking rate

Speech rate and articulation rate values are plotted in Fig. 5 and Fig. 6; Tables 8 – 10 provide the model output and the predicted rate values with confidence intervals, respectively.

Table 8. The fixed-effects output of the *lmer*-model for speech rate.

Predictor	Estimate	Std. error	df	t	p
(Intercept)	**3.85**	**0.079**	**58.664**	**48.592**	**< 0.001**
Age group 2	0.15	0.111	97.927	1.352	0.179
Age group 3	**-0.56**	**0.113**	**73.827**	**-4.945**	**< 0.001**
Gender male	**-0.14**	**0.05**	**5254**	**-2.77**	**0.006**
Age group 2 * Gender male	**-0.48**	**0.073**	**6252.3**	**-6.612**	**< 0.001**
Age group 3 * Gender male	**0.78**	**0.088**	**5444.4**	**8.931**	**< 0.001**

Table 9. The fixed-effects output of the *lmer*-model for articulation rate.

Predictor	Estimate	Std. error	df	t	p
(Intercept)	**5.32**	**0.087**	**56.337**	**61.402**	**< 0.001**
Age group 2	0.23	0.121	79.885	1.865	0.066
Age group 3	**-0.48**	**0.125**	**85.274**	**-3.828**	**< 0.001**
Gender male	-0.03	0.048	7986.9	-0.62	0.535
Age group 2 * Gender male	**-0.42**	**0.069**	**8919.1**	**-6.028**	**< 0.001**
Age group 3 * Gender male	**0.93**	**0.083**	**6965.8**	**11.175**	**< 0.001**

Table 10. Model-predicted speech and articulation rates and 90% confidence intervals in three age groups.

	Gender	Age group 1		Age group 2		Age group 3	
		mean	90% CI	mean	90% CI	mean	90% CI
Sp. rate	Female	3.85	3.72–3.98	4.0	3.87–4.13	3.29	3.15–3.43
	Male	3.71	3.58–3.84	3.38	3.25–3.51	3.93	3.77–4.10
Art. rate	Female	5.32	5.17–5.46	5.54	5.40–5.69	4.84	4.69–4.99
	Male	5.29	5.14–5.43	5.09	4.95–5.24	5.74	5.57–5.91

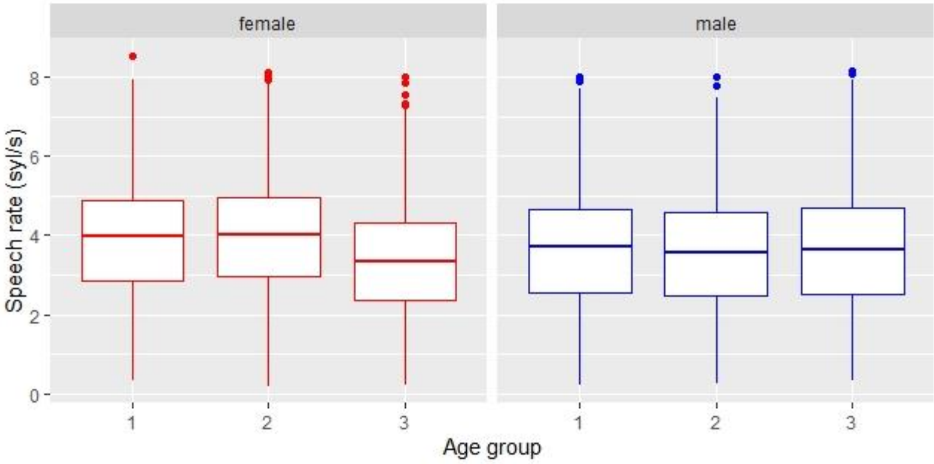

Fig. 5. Speech rates for female and male speakers depending on age.

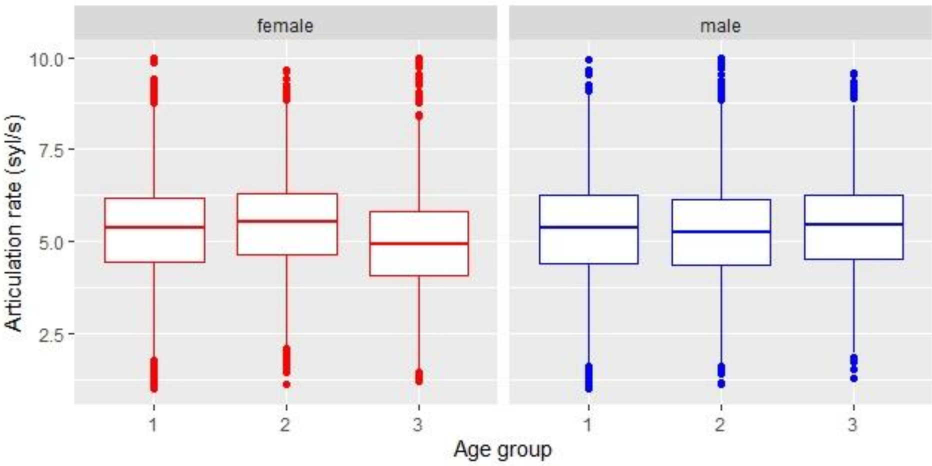

Fig. 6. Articulation rates for female and male speakers depending on age.

4. Discussion

In the paper, the design and development of the Corpus of Estonian Elderly Speech have been introduced and some preliminary analysis results of the prosodic features have been presented. The developed corpus comprises spontaneous speech samples from 200 speakers aged over 60 years and is balanced by gender and age groups. Its size (*ca* 95 hours) is comparable with elderly speech corpora developed for several other languages described in the Introduction. The corpus is aimed to extend the existing Estonian speech resources available for training the acoustic models of Estonian ASR systems. When employed in training, it is expected to further reduce WER in the case of spontaneous and elderly speech, as a literature review of ASR systems for elderly speech confirms (Werner et al. 2019). For example, in the case of a Japanese ASR system WER for elderly speech was reduced from 27.25% to 17.42% after acoustic adaptation using a Japanese corpus of elderly speech (ca 22 hours) (Fukuda et al., 2020). Thus, the introduced Estonian corpus will play a significant role in the development of speech-driven applications targeting the growing elderly population in Estonia (see demographic trends https://ec.europa.eu/eurostat/cache/digpub/demography/bloc-1c.html?lang=en).

Another use of the corpus – research on the acoustic-phonetic characteristics of elderly speech – was performed in the second part of the paper. Besides socio-phonetics, the knowledge about the age-related changes in voice and speech is relevant for distinguishing between normal and pathological speech development.

The age-related changes in the prosodic parameters of Estonian elderly speech explored in the paper are mostly in line with the findings reported in earlier research. For example, Eichhorn et al. (2017) report a significant age-related decrease in F0 in females and a trend of a slight increase in males, whereas several studies report decreasing F0 patterns in both genders or a significant F0 increase in males (see Eichhorn et al., 2017, Table 3). In our corpus, the predicted F0 means showed a significant rising in males (119.23 Hz vs. 124.91 Hz vs. 132.34 Hz in three age groups, respectively) and a slightly decreasing trend in females (189.61 Hz vs. 187.53 Hz vs. 186.03 Hz, respectively). Study-wise differences in F0 development trends can be attributed to the heterogeneous nature of the speech corpora employed in different studies. In addition, it has been shown that age is not the only factor accounting for the changes in F0 characteristics, instead, these are largely determined individually at any age (Markó and Bóna, 2010).

Slowing-down of speaking rate in older age reported in several studies (see e.g., Bóna, 2014 and references therein) is thought to be due to a general slowdown of cognitive and neuromuscular processes and a decline in speech accuracy (Ballard et al., 2001; Ramig, 1983). Our results support previous findings for females: speech rate and articulation rate decrease significantly with age in females (speech rate decreases from 3.85–4.0 syllables per second in age groups 1 and 2 to 3.29 syllables per second in age group 3; articulation rate decreases from 5.32–5.54 syllables per second in age groups 1 and 2 to 4.84 syllables per second in age group 3). Surprisingly, an increasing trend of both speech and articulation rates is observed in males – the rates predicted by a *lmer*-model are higher in the age group 3 (3.93 and 5.75 syllables per second, respectively) than those of age group 1 (3.71 and 5.29 syllables per second, respectively) and age group 2 (3.38 and 5.09 syllables per second, respectively). These rather atypical speaking rates in the oldest male group need further analysis, different grouping by age or involving additional features (e.g., utterance length, education level) into statistical models might help to explain the speaking rate variations.

5. Summary

We have introduced in the paper the development of the Corpus of Estonian Elderly Speech and presented preliminary analysis results of some prosodic characteristics of elderly speakers depending on their age and gender. The corpus will be available for registered users via the Center of Estonian Language Resources (http://keeleressursid.ee/eng/).

6. Acknowledgements

The study has been supported by the European Regional Development Foundation (the project "Centre of Excellence in Estonian Studies") and by the national programme “Estonian Language Technology 2018-2027” (the project “Speech recognition”).

7. References

Albuquerque, L., Oliveira, C., Teixeira, A., Sa-Couto, P., Figueiredo, D. (2019). Age-related changes in European Portuguese vowel acoustics. *Proceedings of Interspeech 2019,* pp. 3965-3969.

Alumäe, T., Tilk, O., Asadullah. (2018). Advanced Rich Transcription System for Estonian Speech. *Frontiers in Artificial Intelligence and Applications,* Human Language Technologies – The Baltic Perspective, pp. 1–8.

Baba, A., Yoshizawa, S., Yamada, M., Lee, A., Shikano, K. (2002). Elderly acoustic models for large vocabulary continuous speech recognition, *Transactions of the Institute of Electronics, Information and Communication Engineers*, D-II J85D-II, pp. 390–397.

Ballard, K. J., Robin, D. A., Woodworth, G., Zimba, L. D. (2001). Age-related changes in motor control during articulator visuomotor tracking, *Journal of Speech, Language, and Hearing Research*, **44**(4), pp. 763–777.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4, *Journal of Statistical Software*, **67**(1), pp. 1–48.

Boersma, P., Weenink, D. (2022). Praat: doing phonetics by computer [Computer Program]. Version 6.2.12, available at http://www.praat.org/

Bóna, J. (2014). Temporal characteristics of speech: The effect of age and speech style, *Journal of the Acoustical Society of America*, **136**(2), EL116-21.

Cucchiarini C., Van hamme, H., van Herwijnen, O., Smits, F. (2006). JASMIN-CGN: Extension of the spoken Dutch corpus with speech of elderly people, children and non-natives in the human-machine interaction modality. *Proceedings of International Conference on Language Resources and Evaluation*, pp. 135–138.

Das, B., Mandal, S., Mitra, P., Basu, A. (2012). Effect of aging on speech features and phoneme recognition: a study on Bengali voicing vowels. *International Journal of Speech Technology,* **16,** pp. 19–31.

Eichhorn, J. T., Kent, R. D., Austin, D., Vorperian H. K. (2017). Effects of Aging on Vocal Fundamental Frequency and Vowel Formants in Men and Women, *Journal of Voice,* **32**(5): 644.e1–644.e9.

Fukuda, M., Nishizaki, H., Iribe, Y., Nishimura, R., Kitaoka, N. (2020). Improving Speech Recognition for the Elderly: A New Corpus of Elderly Japanese Speech and Investigation of Acoustic Modeling for Speech Recognition. *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020),* pp. 6578–6585.

Gorham-Rowan, M. M., Laures-Gore, J. (2006). Acoustic perceptual correlates of voice quality in elderly men and women. *Journal of Communication Disorders,* **39**(3), pp. 171–184.

Hämäläinen, A., Pinto, F. M., Dias, M. S., Júdice, A., Freitas, J., Pires, C. G., Teixeira, V. D., Calado, A., Braga, D. (2012). The First European Portuguese Elderly Speech Corpus. *Proceedings of IberSPEECH*, Madrid, Spain.

Hämäläinen, A., Avelar, J., Rodrigues, S., Dias, M. S., Kolesiński, A., Fegyó, T., Németh, G., Csobánka, P., Ting, K. L. H., Hewson, D. (2014). The EASR Corpora of European Portuguese, French, Hungarian and Polish Elderly Speech. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, May 26-31, Reykjavik, Iceland.

Linville, S. E. (2001). *Vocal aging*, Singular Thomson Learning.

Lippus, P. (2015). *plugin_PhonCorpTools* [Computer program], available at https://gitlab.keeleressursid.ee/partel/plugin_PhonCorpTools/

Markó, A., Bóna, J. (2010). Fundamental frequency patterns: The factors of age and speech type. *Proceedings of the Workshop "Sociophonetics, at the crossroads of speech variation, processing and communication",* pp. 45–48.

Meister, E.; Lasn, J.; Meister, L. (2003). SpeechDat-like Estonian database. *Text, Speech and Dialogue: 6th International Conference*, TSD 2003, Czech Republic, (eds.) Matoušek,V., Mautner, P. Berlin: Springer, pp. 412−417. (Lecture Notes in Artificial Intelligence; 2807).

Meister, E., Meister, L., Metsvahi, R. (2012). New speech corpora at IoC. *Phonetics Symposium 2012,* Tallinn, Estonia, *Proceedings.* (ed.) Meister, E. Tallinn: TUT Press, pp. 30−33.

Meister, L., Meister, E. (2014). Development of the corpus of Estonian Adolescent Speech. *Proceedings of Baltic HLT 2014.* Utka, A., Grigonytė, G., Kapočiūtė-Dzikienė, J., Vaičenonienė, J. (eds.), Amsterdam: IOS Press, pp. 206−209.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K. (2011). *The Kaldi speech recognition toolkit*, in ASRU, 2011.

Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates, *Journal of Communication Disorders,* **16**, pp. 217–226.

RStudio Team 2020. *RStudio: Integrated Development for R*, RStudio, PBC, Boston, MA. Available at http://www.rstudio.com.

Torre P, Barlow J. A. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders,* **42**, pp. 324–333.

Vipperla, R., Renals, S., Frankel, J. (2010). Ageing voices: The effect of changes in voice parameters on ASR performance, *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2010, Article ID 525783, pp. 1–10.

Werner, L., Huang, G., Pitts, B. J. (2019). Automated Speech Recognition Systems and Older Adults: A Literature Review and Synthesis. *Proceedings of the Human Factors and Ergonomics Society 2019 Annual Meeting*.

Xue, S. A., Hao, G. J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production: A pilot study. *Journal of Speech, Language, and Hearing Research*, **46**(3), pp. 689–701.

Received August 17, 2022, accepted August 2, 2022