INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Developmental changes of vowel acoustics in adolescents

Einar Meister, Lya Meister

Tallinn University of Technology, Estonia
einar.meister@taltech.ee, lya.meister@taltech.ee

Abstract

The paper explores the developmental changes of vowel acoustics in Estonian adolescents as a function of age and gender. Formant frequencies F1–F4 and the duration of vowels were measured from read speech samples of 305 native Estonian subjects (173 girls and 132 boys) aged from 10 to 18 years. GAM framework was applied for the statistical analysis. The results show that both the formant frequencies and the vowel space area decrease gradually from 10 to 15 years in both gender groups and the quality of vowels stabilizes at the age of 15–18 years, whereas gender-specific differences emerge around the age of 12–13. Age-related change in the duration of vowels shows similar patterns with formants, however, with no gender difference. The findings are in line with the results reported for adolescent speech in other languages. The analysis results based on speech samples of the subjects with normal linguistic development can be considered reference data for distinguishing between normal and abnormal speech development.

Index Terms: adolescent speech, vowel formants, vowel space area, Estonian

1. Introduction

As repeatedly reported, the acoustic characteristics of children's speech differ widely from those of adult speech, e.g., higher pitch and formant frequencies, and lower speech rate, to name some most striking characteristics of child speech e.g., [1], [2], [3]. The acoustic characteristics of children's speech vary as a function of age and gender due to anatomical and physiological changes. Studies on the human vocal tract have found anatomic gender differences in the oral and pharyngeal areas of the vocal tract in the prepubertal, pubertal and postpubertal age groups (e.g. [4], [5]). Magnetic resonance imaging has revealed that the average length of the vocal tract (as measured from the vocal cords to the outer contour of the lips) is 9.9 cm in children aged 2–4 years, and lengthens to 15.6 cm for boys and 14.4 cm for girls aged 17–18 [6]; the mean length of the adult vocal tract is about 16–17 cm for males and 14.5–15.5 cm for females. Gender differences in the length of the vocal tract are recorded from the age of 12 [5] to 15 [6].

As the acoustic implication of these anatomical changes, age and gender-related differences in fundamental frequency and vowel formants have been documented (e.g., [1], [7], [8]. It has been reported that 7-year-old boys have a mean F0 of 266 Hz and girls of 275 Hz, and the gender difference becomes significant at age 12, where boys have a mean F0 of 226 Hz and girls 231 Hz; the abrupt change in the mean value of boys F0 occurs between the ages of 12 and 15, falling by ca 100 Hz by the age of 15, later changes are marginal [1]. Vowel formant frequencies of boys are reported systematically lower than those of girls from age 7–8 years; on average, gender difference appears to be statistically significant at age 12 and becomes fully distinguishable around age 15, as reviewed in [7]. The decrease in formant frequencies during childhood and adolescence is accompanied by a decrease of the vowel space area (VSA) till age 15–16. However, data reported on gender differences in VSA vary widely, e.g., according to [9], a gender difference in VSA appears as early as age 4, while data from [10] show that it does not appear until age 16.

In parallel with anatomic changes in childhood, the development of the speech motor system takes place which is manifested in the increase of speech and articulation rates. According to various studies, the articulation rate in native English children with typical linguistic development aged 3–6 years varies from 2.9 to 4.3, and in children aged 7–12 years from 4.5 to 5.6 syllables per second [11]. The speech rate increases with age equally for boys and girls, and there are no gender differences within age groups. Speech tempo directly affects the duration of different speech segments – the faster the tempo, the more speech segments are produced in a time unit and the shorter the duration of speech segments. For example, the duration of adult Estonian vowels in slow speech can be about two times longer than that of fast speech [12].

The paper investigates the acoustic characteristics of Estonian vowels of adolescent speech. The acoustics of vowels produced by adult speakers have been explored in a considerable number of studies (see [13] for an overview]), yet the developmental characteristics of Estonian vowels from childhood to adults remain to be discovered. Some preliminary results on age- and gender-specific developments of fundamental and formant frequencies, and speech tempo have been presented in [14], [15], and [16]. This paper aims to expand the knowledge on the spectral and temporal features of vowels in adolescent speech as a function of age and gender.

The Estonian vowel system includes nine vowels /i ü u e ö õ o ä a/ characterized by the articulatory features as shown in Table 1.

Table 1. *Articulatory features of Estonian vowels.*

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
High	/i/ [i]	/ü/ [y]	/õ/ [ɯ]	/u/ [u]
Mid	/e/ [e]	/ö/ [ø]	[ɘ] [ɤ]	/o/ [o]
Low	/ä/ [æ]		/a/ [ɑ]	

All vowels occur in a primary stressed syllable, whereas only five vowels [ɑ e i o u] occur in non-initial syllables; there is no vowel harmony in Estonian (unlike the related languages Finnish and Hungarian) [17]. Duration of vowels is mostly guided by the three-way quantity system of Estonian involving contrastive prosodic patterns traditionally referred to as short (Q1), long (Q2), and overlong (Q3) quantity degrees [18] (see

http://dx.doi.org/10.21437/Interspeech.2021-1649

[13] for an overview of the Estonian quantity system). Also, the intrinsic duration differences of vowels have been reported in acoustic measurements [19] and verified in a perceptual study [20]. In vowels, the three-way length contrast occurs in the primary stressed first syllables, and there is no length contrast in unstressed syllables (vowels in unstressed syllables are phonologically short, although their duration may vary to a great extent). In read speech, the quality differences of stressed vowels in different length contrasts do not exceed 1 Bark in the F1-F2 vowel space, but vowel quality in unstressed syllables is reduced [21]. However, in spontaneous speech, vowel quality was significant for the contrast of short and long categories [22].

The research questions of the current paper focus on the developmental changes of vowel formants and duration depending on gender and age. We hypothesize that: (1) formant values of stressed and unstressed vowels, vowel space area, and vowel durations in Estonian adolescent speech exhibit similar developmental patterns reported earlier for other languages, (2) gender-specific differences in formant values emerge at the age of 10–13 and become more salient in later ages, (3) vowel durations shorten equally with age in boys and girls, there are no gender differences within age groups, (4) in all age groups, the acoustic characteristics of vowels consistently reveal the patterns typical to the native language phonological categories.

2. Method

2.1. Speech material

The Estonian Adolescent Speech Corpus [14] consisting of speech samples read by 309 subjects (175 girls and 134 boys) in the age range of 9 to 18 years was used in the acoustic analysis. The corpus represents cross-sectional speech data of different age groups recorded in ten schools around Estonia. The subjects were recruited according to given criteria (native Estonian, no hearing and speaking disorders, fluency in reading of unfamiliar texts) by the school teachers, approvals from the headteacher and parents were obtained beforehand. The recordings were performed using a mobile recording set including a laptop with BAS SpeechRecorder software [23], a microphone preamplifier M-Audio Mobile Pre, two microphones (desktop microphone Audio-Technica ATM33a and close-talking microphone Sennheiser ME3), and an external monitor to show the prompts. The signals were stored directly on the hard disc in wav format (sampling at 44.1 kHz, resolution 16 bits). 70 items (ca 15 minutes of speech) per speaker were recorded, resulting in ca 70 hours of speech in total. The corpus is available for registered users via the Center of Estonian Language Resources and internationally via the EU CLARIN infrastructure.

The corpus contains linguistically diverse material – digits, phone numbers, time expressions, IT terms, sentences with name entities, phonetically rich sentences, and several samples of spontaneous speech elicited with pictures to be described and with topics for storytelling. However, in the current study, we have used a subset of the corpus consisting of 21 phonetically rich sentences per subject read by 173 girls and 132 boys in the age range from 10 to 18 years (9-years-old subjects were excluded due to sparse data as there were only 4 subjects in this age group). The utterances (in total 6405) were segmented manually on the word and phone levels.

2.2. Acoustic analysis

Formant frequencies F1–F4 and the duration of vowels were measured by using a Praat [24] script that implements the Burg method with adapted parameters for different gender and age groups. Formant ceiling of 5500 Hz was applied for all girls as well as for boys aged 10–13, and of 5000 Hz for boys aged 14–18; max number of formants was 5, window length 0.025 s, pre-emphasis 50 Hz. The formant frequencies on 65720 vowels were measured from a section around the vowel midpoint that accounted for 20% of the total vowel duration. To eliminate obvious outliers, mean values and standard deviations of F1 – F4 were calculated for each age group and those formant values that deviated by more than ±1,5 standard deviations from the mean were excluded from further analysis. The final data set consisted of 60216 vowels grouped according to their position in a word – first syllable vowels (in total 24502, hereinafter referred to as V1) and the vowels in further syllables (in total 35714, hereinafter referred to as V2) (Table 2). As a rule, V1 vowels are stressed and V2 vowels unstressed.

Table 2. *The number of analyzed stressed (V1) and unstressed (V2) vowels.*

	[i]	[y]	[u]	[e]	[ø]	[ɤ]	[o]	[æ]	[ɑ]
V1	2590	1276	2792	3082	604	1096	4010	1536	7516
V2	9361	-	4692	8793	-	-	639	-	12229

To explore the change of vowel formants over the age range from 10 to 18, the formant dynamics method ([25]) was adapted. According to the method, formant values are measured at regular time intervals throughout the vowel and are presented as a trajectory on the F1-F2 acoustic space describing the change of vocal quality. In our case, the time interval is one year and the measured values are the means of F1 and F2 for each age group by gender. Thus, the vowel trajectory consists of eight sections (10–11, 11–12, 12–13, 13–14, 14–15, 15–16, 16–17, and 17–18 years), whereas the vowel section length (VSL) is given by the formula:

$$VSL_n = \sqrt{(F1_n - F1_{n+1})^2 + (F2_n - F2_{n+1})^2} \tag{1}$$

The vowel trajectory length (VTL) is found as the sum of the eight sections:

$$VTL = \sum_{n=1}^{8} VSL_n \tag{2}$$

The vowel space area (VSA) is another measure exploited to explore the developmental changes of vowel acoustics in different subject groups and to perform cross-linguistic comparison of vowel systems (e.g., [26], [27]). We calculate the VSA for each age group by gender as the area between the corner vowels using the formula from [27]:

$$VSA = 0.5\,\{(\mathrm{F2/i/{*}F1/æ/} + \mathrm{F2/æ/{*}F1/a/} + \mathrm{F2/a/{*}F1/u/} + \mathrm{F2/u/{*}F1/i/}) - (\mathrm{F1/i/{*}F2/æ/} + \mathrm{F1/æ/{*}F2/a/} + \mathrm{F1/a/{*}F2/u/} + \mathrm{F1/u/{*}F2/i/})\} \tag{3}$$

3. Results

The General Additive Mixed Model (GAMM) framework was applied to estimate the relationships between the predictor variables (age, gender, quantity) and the response (F1–F4, VTL, VSA, duration) using *mgcv* package [28] within RStudio environment [29]. To validate the models and visualize the results *itsadug* package was applied [30].

3.1. Formant frequencies

GAMMs were fitted to predict F1–F4 values of all nine stressed vowels (V1) with the smooth term *age* interacting with factor *gender*, factors *vowel* and *quantity*, and *subject* as the random effect; in GAMMs for unstressed vowels (V2) *quantity* was excluded as there is no quantity contrast in V2. The models for F1–F4 showed significant main effects of all predictors in both stressed and unstressed vowels. The plots of partial effects of the explanatory features on F1 and F2 are shown in Figure 1 for stressed and in Figure 2 for unstressed vowels. (The numeric summaries of all models and the plots for F3 and F4 are omitted due to space savings).

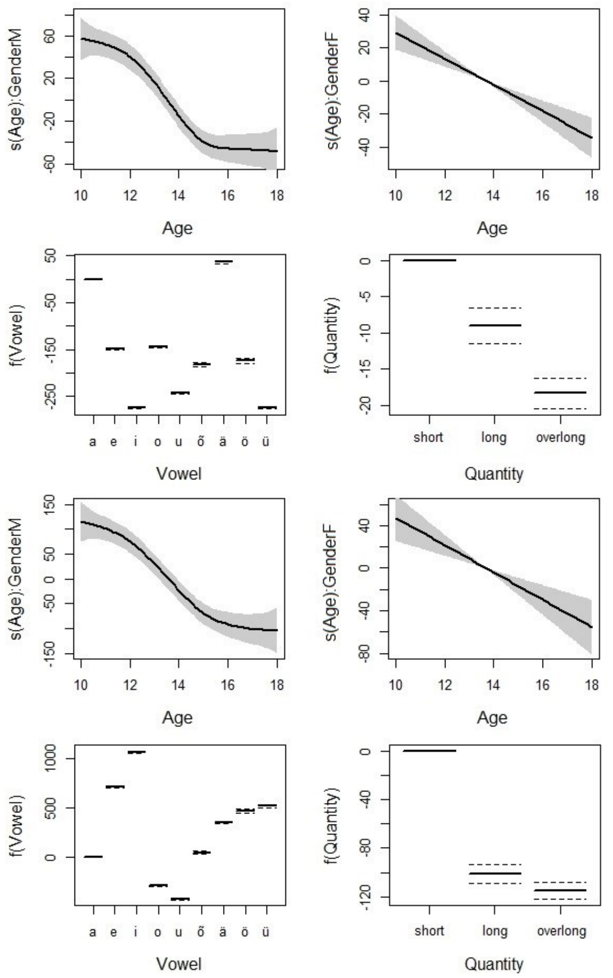

Figure 1: *Response plots of the GAMMs for F1 (2 top rows) vs. F2 (2 bottom rows) of the stressed vowels (y-axis represents the centered values of F1 and F2).*

Age-related changes of F1–F4 show all decrease with age, however, with slightly different patterns for males and females. The formant values in girls are always higher than in boys. The post-hoc Wald test revealed that in most vowels the decrease of formant values and gender-specific differences are not significant till age 12, larger changes ($p < 0.05$–0.001) occur between the ages of 12 and 15, with boys' formants declining faster than that of girls. At the age of 16–18, the changes in formants are small and within that period the vowels acquire a stable quality, preserving the individual variability of the subjects. However, in most vowels, the mean formant values at age 18 have remained somewhat different than that of adult males and females reported in several other studies (cf. e.g., [13]). It suggests that the rather stable vowel quality achieved by age 16–18 continues to evolve during a longer period of life.

The length category of vowels is significant in F1–F3 in all contrasts ($p < 0.001$), in F4 for the short vs. overlong contrast only ($p < 0.05$).

The quality of the unstressed (V2) vowels is always reduced compared to the stressed (V1) counterparts resulting in a more central location in the F1-F2 vowel space. As expected, the most reduced vowel in adolescent speech is /e/.

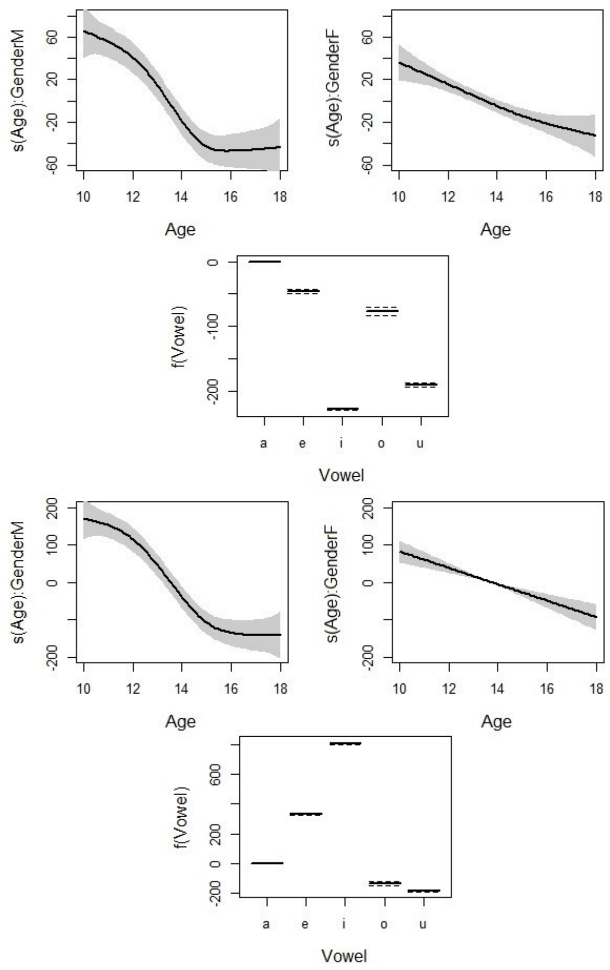

Figure 2: *Response plots of the GAMMs for F1 (2 top rows) vs. F2 (2 bottom rows) of the unstressed vowels (y-axis represents the centered values of F1 and F2).*

3.2. Vowel trajectory length (VTL)

The VTL values for male and female subjects are given in Table 3 and plotted in Figure 3 to represent the development of vowel quality from 10 to 18 years in one-year steps.

Table 3. *VTL values (in Hz) for stressed (V1) vowels.*

	[i]	[y]	[u]	[e]	[ø]	[ɤ]	[o]	[æ]	[ɑ]
M	744	470	94	592	418	245	182	504	328
F	366	273	73	347	220	155	146	297	169

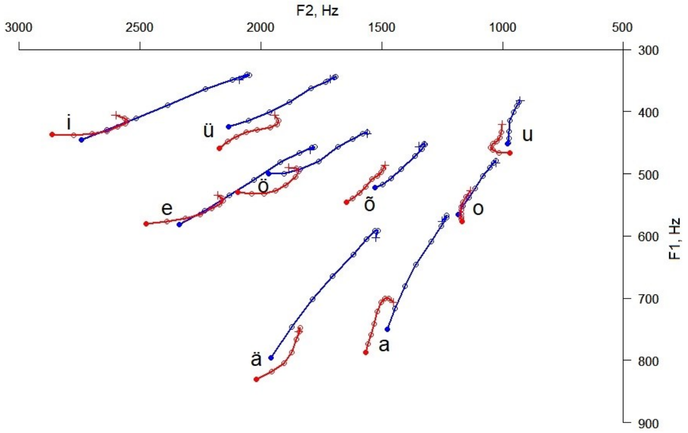

Figure 3: *Vowel trajectories for male (blue) and female (red) subjects represent the development of vowel quality from 10 to 18 years. Filled circles represent the vowel locations at age 10, crosses denote the vowel positions at age 18, empty circles indicate the one-year steps.*

3.3. Vowel space area (VSA)

VSA follows a similar pattern with formant development revealing a consistent decrease till age 15 and followed by a rather stable size in 16–18 years. The plots in Figure 4 illustrate the location of the corner vowels for males and females and the VSA boundaries of 10-years-old and 18-years-old groups.

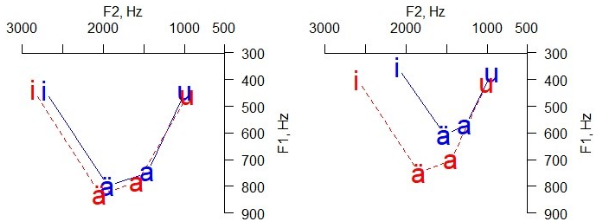

Figure 4: *VSA plots for 10-year-old (left) and 18-year-old (right) males (blue) and females (red).*

3.4. Vowel duration

GAMM for the duration of stressed vowels showed a significant effect of age and quantity ($p < 0.001$), but not gender ($p = 0.94$). Vowel duration decreases till age 15 and remains stable in later ages whereas the duration ratios between long vs. short 1.85 and overlong vs. short 2.1 are constant in all age groups.

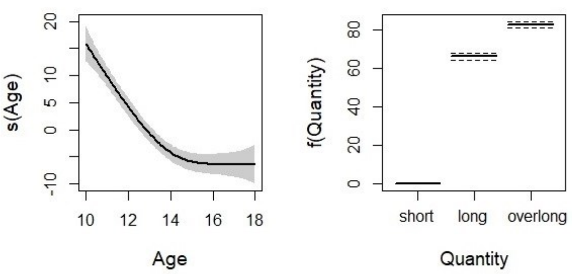

Figure 5: *Response plots of the GAMM for duration of the stressed vowels.*

4. Discussion

The paper reports the results on the acoustic characteristics of vowels depending on the subjects' age and gender. The results show that both the formant frequencies and the vowel space area decrease gradually from 10 to 15 years in both gender groups, and the quality of vowels stabilizes at the age of 16–18 years, whereas gender-specific differences emerge around age 12–13. These findings are in line with studies addressing the anatomical development of the vocal tract. The quality change measured as VTL is largest in the front vowels /i/ and /e/, and smallest in the back vowels /u/ and /o/.

Like in adult speech, short stressed vowels are produced more centrally than long and overlong vowels, and the unstressed vowels are qualitatively reduced. Age-related change in the duration of vowels shows similar patterns with formants decreasing till age 15 and remaining stable in ages 16–18. Interestingly, the duration ratios between different quantity contrasts are constant in all age groups and do not exhibit gender differences. The long vs. short duration ratio of 1.85 is close to the corresponding adult ratio of 1.9, but the overlong vs. short ratio of 2.1 is smaller than the typical ratio of adult speakers 2.5 [31]. The stabilization of vowel durations at age 15 might be related to the maturation of the speech motor system.

5. Conclusions

This study examined the developmental changes of vowel acoustics in Estonian adolescents as a function of age and gender. The findings are in line with the results reported for adolescent speech in other languages and reveal the patterns of the native language phonology. The reported results can serve as reference data that are typical for Estonian-speaking subjects aged from 10 to 18 with normal linguistic development.

6. Acknowledgements

The study has been supported by the European Regional Development Foundation (the project "Centre of Excellence in Estonian Studies").

7. References

[1] S. Lee, A. Potamianos, and S. Narayanan, "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," The Journal of the Acoustical Society of America, vol. 105, no. 3, pp. 1455–1468, 1999.

[2] J. E. Huber, E. T. Stathopoulos, G. M. Curione, T. A. Ash and K. Johnson, "Formants of children, women, and men: The effects of vocal intensity variation," *Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1532–1542, 1999.

[3] E. Jacewicz, R. A. Fox and L. Wei, "Between-speaker and within-speaker variation in speech tempo of American English", *Journal of the Acoustical Society of America,* vol 123, no. 2, pp.839–50, 2010.

[4] L. Ménard, J.-L. Schwartz, L.-J. Boë, J. Aubin, "Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model," *Journal of Phonetics,* vol. 35, pp. 1–19, 2007.

[5] H. K. Vorperian, S. Wang, D. Schimek, B. Reid, R. D. Kent, L. R. Gentry and M. K. Chung, "Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study," *Journal of Speech, Language, and Hearing Research*, vol. 54, no. 4, pp. 995–1010, 2011.

[6] T. W. Fitch and J. N. Giedd, "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," *Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1511–1522, 1999.

[7] H. K. Vorperian and R. D. Kent, "Vowel acoustic space development in children: A synthesis of acoustic and anatomic data," Journal of Speech, Language, and Hearing Research, vol. 50, no. 6, pp. 1510–1545, 2007.

[8] S. P. Whiteside, "Sex-specific fundamental and formant frequency patterns in a cross-sectional study," Journal of the Acoustical Society of America, vol. 110, no. 1, pp. 464–478, 2001.

[9] T. L. Perry, R. N. Ohde and D. H. Ashmead, "The acoustic bases for gender identification from children's voices," *Journal of the Acoustical Society of America*, vol. 109, no. 6, pp. 2988–2998, 2001.

[10] P. J. R. Flipsen and S. Lee, "Reference data for the American English acoustic vowel space. – Clinical Linguistics & Phonetics, vol. 26, no. 11–12, pp. 926–933, 2012.

[11] K. J. Logan, C. T. Byrd, E. M. Mazzocchi and R. B. Gillam, "Speaking rate characteristics of elementary-school-aged children who do and do not stutter," *Journal of Communication Disorders*, vol. 44, pp. 130–147, 2011.

[12] A. Eek and E. Meister, "Foneetilisi katseid ja arutlusi kvantiteedi alalt (I): häälikukestusi muutvad kontekstid ja välde," *Keel ja Kirjandus*, vol. 46, no. 11, pp. 815−837, 2003.

[13] E. L. Asu, P. Lippus, K. Pajusalu and P. Teras, *Eesti keele hääldus.* Tartu: Tartu Ülikooli Kirjastus, 2016.

[14] E. Meister and L. Meister, "Eesti laste kõne I. Põhitooni akustiline analüüs" [Estonian children's speech I. Acoustic analysis of fundamental frequency], *Keel ja Kirjandus*, no. 7, pp. 518–533, 2017.

[15] E. Meister and L. Meister, "Eesti laste kõne II. Vokaalide akustiline analüüs," [Estonian children's speech II. Acoustic analysis of vowels], Keel *ja Kirjandus*, no. 4, pp. 282–295, 2019.

[16] E. Meister and L. Meister, "Eesti laste kõne III. Kõnetempo ja silbikestuste analüüs," [Estonian children's speech III. Analysis of speech rate and syllable durations], Keel *ja Kirjandus*, 2021 (to appear).

[17] E. L. Asu and P. Teras, "Illustrations of the IPA: Estonian," *Journal of the International Phonetic Association*, vol. 39, no. 3, pp. 367−372, 2009.

[18] Lehiste, I. 1960. Segmental and syllabic quantity in Estonian. American Studies in Uralic Linguistics 1, Bloomington: Indiana University Press, 21-82.

[19] E. Meister and S. Werner, "Intrinsic microprosodic variations in Estonian and Finnish: Acoustic analysis," in Aulanko, R., Wahlberg, L., Vainio, M. (eds.), *The Phonetics Symposium 2006,* Department of Speech Sciences, University of Helsinki, vol. 53, pp. 103–112, 2006.

[20] E. Meister and S. Werner, "Duration affects vowel perception in Estonian and Finnish," *Linguistica Uralica*, vol. 45, no. 3, pp. 161–177, 2009.

[21] A. Eek and E. Meister, "Quality of Standard Estonian vowels in stressed and unstressed syllables of the feet in three distinctive quantity degrees", *Linguistica Uralica*, vol. 34, no. 3, pp. 226–233, 1998.

[22] P. Lippus, E. L. Asu, P. Teras and T. Tuisk, "Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian," *Journal of Phonetics*, vol. 41, no. 1, pp. 17−28, 2013.

[23] Chr. Draxler and K. Jänsch. "SpeechRecorder - a Universal Platform Independent Multi-Channel Audio Recording Software", in Proc. of LREC, pp. 559−562, Lisbon, 2004.

[24] P. Boersma and D. Weenink, *Praat: doing phonetics by computer* [Computer program]. Version 6.1.41, retrieved 25 March 2021 from http://www.praat.org/

[25] R. A. Fox and E. Jacewicz, "Cross-dialectal variation in formant dynamics of American English vowels," *Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2603–2618, 2009.

[26] H. Chung, E. J. Kong, J. Edwards, G. Weismer, M. Fourakis and Y. Hwang, Cross-linguistic studies of children's and adults' vowel spaces," *Journal of the Acoustical Society of America*, vol. 131, no. 1, pp. 442–454, 2012.

[27] P. J. R. Flipsen and S. Lee, "Reference data for the American English acoustic vowel space," *Clinical Linguistics & Phonetics*, vol. 26, no. 11–12, pp. 926–933, 2012.

[28] S. N. Wood, *Generalized Additive Models: An Introduction with R (2nd ed.),* CRC Press, 2017.

[29] RStudio Team, *RStudio: Integrated Development for R,* RStudio, PBC, Boston, MA URL http://www.rstudio.com/, 2020.

[30] J. Van Rij, M. Wieling, Martijn, H. Baayen and H. van Rijn, https://cran.r-project.org/web/packages/itsadug/, 2020.

[31] L. Meister, The Perception and Production of Estonian Vowel and Duration-Based categories by Non-Native Subjects with a Russian-Language Background. An Experimental Phonetic Study, [in Estonian], PhD thesis, Tartu University, 2011.