https://doi.org/10.7592/FEJF2023.90.meister_meister

DEVELOPMENTAL CHANGES IN ACOUSTIC CHARACTERISTICS OF SPEECH OF ESTONIAN ADOLESCENTS

Einar Meister
Senior Research Fellow
Tallinn University of Technology, Estonia
einar.meister@ttu.ee

Lya Meister
Research Fellow
Tallinn University of Technology, Estonia
lya.meister@ttu.ee

Abstract: The paper introduces the Estonian Adolescent Speech Corpus and explores the developmental changes in speech production based on acoustic characteristics of fundamental frequency (F0), formant frequencies, and speech tempo as a function of age and gender. Age- and gender-related anatomic changes in adolescence have implications for speech acoustics: a sudden drop of F0 at puberty in boys, and an almost gradual decrease of the acoustic vowels space. In parallel with anatomic changes, the development of the speech motor system is manifested as the increase of speaking rate. The analysis of fundamental frequency (F0) shows that in both male and female speakers, the F0 decreases gradually at the age from 9 to 12 years, then in males F0 drops ca by 100 Hz at the age of 12–15 due to puberty voice change, and becomes stable at the age of 15–18; in female speakers, a gradual decrease of F0 continues till the age 18. The formant frequencies of vowels decrease gradually from 10 to 15 years in both genders and the quality of vowels stabilizes at the age of 15–18 years, gender-specific differences emerge at the age of 12–13. Speech rate increases from 4 syllables per second in 9–10 years to 5.1 syllables per second in 14 years and becomes stable between the ages of 15 and 18, gender differences are not significant. The results of the current study can be considered as reference data that are typical for Estonian-speaking individuals aged 9–18 years with normal language development.

Keywords: acoustic analysis, adolescent speech, formant frequencies, fundamental frequency, speech corpus, speech motor development, speech tempo

INTRODUCTION

The acoustic characteristics of children's speech differ considerably from those of adults' speech, e.g., higher pitch and formant frequencies, longer segmental durations, and lower speech rate are repeatedly reported features of children's speech (e.g. Lee et al. 1999; Jacewicz et al. 2010). Age- and gender-related anatomic changes of the vocal apparatus during childhood and adolescence are manifested in a sudden drop of fundamental frequency (F0) at puberty in boys, and in a gradual decrease of the acoustic vowels space and formant frequencies in both genders. In parallel with anatomic changes, the development of speech-motor control as well as cognitive and linguistic processing takes place, which is revealed in the increase of speech and articulation rates (Smith & Zelaznik 2004; Logan et al. 2011).

The length of the vocal folds in infants (younger than 1 year) is 4-5 mm, at the age of 20, the vocal folds have reached a length of approx. 11-15 mm in women and 17-25 mm in men (Hirano et al. 1981; Rogers et al. 2014). The vocal fold growth and the enlargement of the larynx have acoustic implications to F0, and consequently on the perceived voice pitch. According to Lee et al. (1999), the mean F0 for 7-year-old English-speaking boys is 266 Hz and for girls 275 Hz (the difference is not statistically significant), the gender difference becomes significant at the age of 12, where the mean F0 for boys is 226 Hz and for girls 231 Hz. A large change in F0 for boys occurs between the ages of 12 and 15, dropping to 127 Hz by age 15, with marginal changes thereafter. A study of German youth aged from 13 to 19 found that boys' mean F0 dropped by about 80 Hz (206.7 Hz – 126.6 Hz) between ages 13–15 and only 4 Hz between ages 15–19 while girls' F0 fell steadily from 230 Hz to 218 Hz between the ages of 13 and 19 (Draxler et al. 2008).

During speech production, the shape of the vocal tract varies depending on the position of the jaw, tongue, and lips, and as a result, forming different spectral and temporal patterns for different speech segments. The acoustic quality of vowels is primarily determined by the first two formants (F1, F2), which form the acoustic vowel space (Fant 1960; Stevens 2000). The formant frequencies of vowels and the size of the vowel space are directly related to the anatomical size of the vocal tract – a longer vocal tract results in lower formant frequencies compared to a shorter vocal tract. As found in magnetic resonance imaging studies, the average length of the vocal tract (measured from the vocal folds to the lips) is 9.9 cm in 2–4-year-old children, and 13.9 cm in 13–14-year-old children. Differences in the length of the vocal tract of boys and girls have been recorded since the age of 15: 14.6 cm in 15–16-year-old boys and 13.7 cm in girls, 15.6 cm in 17–18-year-old boys and 14.4 cm in girls (Fitch & Giedd 1999). In a work aggregating data from different studies (Vorperian & Kent 2007), it has been found that the formant frequencies of boys' vowels are systematically lower than the formant frequencies of girls from the age of 7–8, the gender difference becomes statistically significant at the age of 12 and develops further till the age of 15.

Articulatory movements involve temporal and spatial control, which requires the coordinated interaction of the motor and the language systems (Smith 2006). Research on speech-motor development (e.g. Sharkey & Folkins 1985; Smith & Goffman 1998; Goffman & Smith 1999; Green et al. 2000; Schötz et al. 2013; Barbier et al. 2020) has shown that children's articulatory movements are slower and more variable than those of adults. For example, a study by Smith & Zelaznik (2004) investigated native English-speaking children and adults (in total 180 subjects) aged 4–22 years by recording upper lip, lower lip, and jaw movements while reading different sentences. The results showed that with increasing age, the variability of articulatory movements decreases and the time taken to form a sentence shortens, and the processes of speech-motor control in both boys and girls become similar to adults only after the age of 14. The authors conclude that the age-related increase in speaking rate is due to improvements in cognitive and linguistic processing and speech-motor control.

The most common measures of speaking rate – speech rate and articulation rate – are calculated as the number of speech units (words, syllables, or segments) produced in a unit of time (a minute or second) (Tsao et al. 2006). Speech rate includes pauses in the utterance (e.g. hesitations, pauses between words), while calculating the articulation rate, pauses longer than 250 ms are excluded (e.g. Ingham & Riley 1998; Crystal & House 1990). According to various studies, the articulation rate in the spontaneous speech of 3–6-year-old native English-speaking children varies from 2.9 to 4.3, and the speech rate ranges from 2.3 to 2.6 syllables per second, for 7–12-year-olds, 4.5 to 5.6 and 2.4 to 2.9 syllables per second, respectively (Logan et al. 2011: Table 1). As a rule, the speaking rate increases equally with age in boys and girls, and the differences within age groups are not statistically significant (mostly, the speaking rate of male speakers is slightly faster) (e.g. Robb et al. 2004; Verhoeven et al. 2004; Jacewicz et al. 2009, 2010; Lee & Doherty 2017). In adults, speaking rates of 220–280 syllables per minute (3.7–4.7 syllables per second) and articulation rates of 200–346 syllables per minute (3.3–5.8 syllables per second) have been documented (Lee & Doherty 2017: Table 1).

While reading aloud written texts, the speaking rate is affected by the text length as confirmed in several studies (e.g. Lehiste 1974; Sadagopan & Smith 2008; Amir & Grinfeld 2011; Bishop & Kim 2018; Darling-White & Banks 2021). This is known as anticipatory shortening, according to which in the speech planning process the speaker adjusts his average syllable duration to the expected length of the phrase (see, e.g., Bishop & Kim 2018 and references therein). Sadagopan and Smith (2008) investigated the relationship between text length and speaking rate in a comparison of children (aged 5–16) and adults (aged 20–23). Their study revealed that the duration of the test phrase read in isolation is longer than when read in a frame sentence both in adults and in 9–16-year-olds, but not in 5–7-year-old children. The authors hypothesize that children and adults use different motor planning strategies when reading more complex sentences: younger children plan their speech in smaller speech units (words or syllables) while older children and adults in longer speech units (phrases). The results also suggest that the transition to an adult-like speech-motor planning strategy begins around age 9.

The present study aims to document the acoustic characteristics of Estonian adolescent speech and gain a better understanding of the variations related to the speaker's age and gender. In particular, we will explore (1) the acoustic variations of F0, (2) vowel formants and duration, and (3) changes in speech tempo using the measures of speaking rate.

MATERIALS AND METHODS

Speech corpus

The Estonian Adolescent Speech Corpus (Meister & Meister 2014) consists of speech samples from 309 subjects (175 girls and 134 boys) in the age range from 9 to 18 years (Table 1). The corpus represents cross-sectional speech data of different age groups. The subjects were recruited in ten schools across Estonia (four schools in the capital area, two in the North-East, two in the South-East, and two on the island Saaremaa in Western Estonia). The school teachers selected the volunteers according to the given criteria – native Estonian, no hearing and speaking disorders, and fluency in the reading of unfamiliar texts. All subjects signed a consent form and filled out a questionnaire in which they provided information about their age, gender, place of residence, class, school, mother tongue, and foreign language learning. Written consent was also obtained from the parents and the schools.

Table 1. *The distribution of subjects by age and gender.*

Age	Male	Female	Total
9	2	2	4
10	12	12	24
11	18	22	40
12	12	23	35
13	18	32	50

14	17	28	45
15	21	15	36
16	16	20	36
17	8	11	19
18	10	10	20
Total	**134**	**175**	**309**

For the recordings, a text corpus was compiled that contained linguistically diverse material: phonetically rich sentences, sentences containing names of places, persons, and organizations, time expressions, phone numbers, random number sequences, IT terms, and short stories. To elicit spontaneous speech, the subjects were asked to describe pictures and talk about themselves, their family, school, friends, and hobbies or tell a story on a freely chosen topic. From each speaker, 60 read and 10 spontaneous items were recorded in a quiet room using a laptop with BAS SpeechRecorder software (Draxler & Jänsch 2004), two microphones (desktop and close-talking microphone), 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). In total, the corpus contains approximately 70 hours of speech, about 15 minutes from each subject.

In the current study, the acoustic analyses were performed on a subcorpus of read speech samples consisting of 21 phonetically rich sentences per subject, in total 6489 read utterances. The duration of utterances ranged from 4.95 to 13.67 seconds, with an average of 7.8 seconds. All utterances were segmented manually on the word and phone levels using Praat (Boersma & Weenink 2022). Syllable boundaries and types were added using a custom Praat script (Lippus 2015).

Acoustic analysis

In the study, the following acoustic features were investigated: (1) fundamental frequency (F0), (2) vowel formants F1 and F2, and vowel duration, and (3) speech and articulation rates.

F0

For the F0 analysis, a custom Praat script was compiled using the two-step procedure recommended by Hirst (2007). First, the F0 values of each utterance were found in the frequency range 75–600 Hz, then the range was narrowed according to the speaker's F0 variation as follows: F0 max = 1.5 x 3rd quartile value, F0min = 0.75 x 1st quartile value. This approach takes into account the subject's individual F0 range and thus provides more reliable results. However, about 20% of the utterances required manual F0 validation in cases where the automatically found F0 values seemed unlikely, e.g. when boys had maximum F0 values above 400 Hz or girls had minimum F0 values below 150 Hz. There were also more errors in the speech of boys with a voice mutation period, where F0 variations were larger and sometimes reached the falsetto register.

Vowel-related features

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 2.

Table 2. *Articulatory features of Estonian vowels.*

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

All vowels occur in a primary stressed syllable and only five vowels [ɑ e i o u] can occur in non-initial syllables. The duration of vowels is mainly defined by the three-way quantity system of Estonian (Lehiste 1960; see Asu et al. 2016 for further references). The three-way length contrast in vowels occurs in the primary stressed first syllables of a foot, and there is no length contrast in the unstressed syllables.

The formant frequencies F1–F2 and the duration of vowels were measured using a custom Praat script that implements the Burg method with adapted parameters for different gender and age groups. For all girls and boys aged 10–13 the formant ceiling value of 5500 Hz was applied, and of 5000 Hz for boys aged 14–18; the max number of formants was 5, window length 0.025 s, pre-emphasis 50 Hz. The formant frequencies of 65720 vowels were measured around the vowel midpoint. The formant values were obtained first in the Hz scale and then converted to the psychoacoustic Bark scale. Mean values and standard deviations of F1 and F2 were calculated for each age and gender group and the values that deviated by more than ±1.5 standard deviations from the group mean (as obvious outliers) were excluded from further analysis. The final data set consisted of 60 216 vowels grouped according to their position in a word – first (stressed) syllable vowels (in total 24500, hereinafter referred to as V1) and the vowels in second (unstressed) syllables (in total 23 343, hereinafter referred to as V2) (Table 3). The vowels in further syllables (V3–V5) were not included in the analysis. Formant data of 9-year-old subjects were excluded as this group included 2 boys and 2 girls only.

Table 3. *The number of analyzed stressed (V1) and unstressed (V2) vowels.*

	[ɑ]	[e]	[i]	[o]	[u]	[ɤ]	[æ]	[ø]	[y]
V1	7515	3081	2590	4010	2792	1096	1536	604	1276
V2	7437	4930	7426	262	3288	-	-	-	-

In order to explore the change of vowel quality over the age range from 10 to 18, the means of F1 and F2 of all vowel categories for each age and gender group were calculated and presented as a developmental trajectory in the F1 by F2 acoustic plane. The developmental vowel trajectory consists of eight sections corresponding to each consecutive age group (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 (Fox & Jacewicz 2009):

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

The developmental vowel trajectory length (DVTL) is the sum of the eight sections:

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

Speech tempo

Two measures of speech tempo were used, the speech rate and the articulation rate. To derive each measure, the number of syllables in each utterance was counted, and utterance duration was measured. Pauses shorter than 250 ms were included in the calculation of speech rate and were excluded when calculating articulation rate. Utterances with up to 20 syllables were used for further analysis (in total 7125, the average number of syllables in utterance 11.8, median 11). Two speaking rate measures for each utterance were calculated:

- the speech rate = the number of syllables in an utterance / the duration of an utterance including pauses,
- the articulation rate = the number of syllables in an utterance / the duration of an utterance without pauses.

Statistical analysis

The R environment (R Core Team 2018) within the RStudio program (RStudio Team 2020) was used for statistical data processing. Generalized Additive Mixed Models (GAMM) with the *mgcv* package (Wood 2017) were used for modeling, *itsadug* package (van Rij et al. 2022) was used for model validation and visualization of the results; for formant plots, we used the R package *phonR* (McCloy 2016).

RESULTS

F0

The histograms (Figure 1) show the distribution of mean F0 in boys and girls. In the case of boys, the F0 distribution is binomial with clearly distinguished peaks at 110 Hz and 215 Hz. The first peak represents the most frequent mean F0 value of boys who have undergone a pubertal voice change period (boys aged 15–18), and the second peak corresponds to the most frequent mean F0 value of younger boys (aged 9–12) before voice change. As expected, there is only one peak in the histogram for girls, representing the median F0 value for all girls (226 Hz).

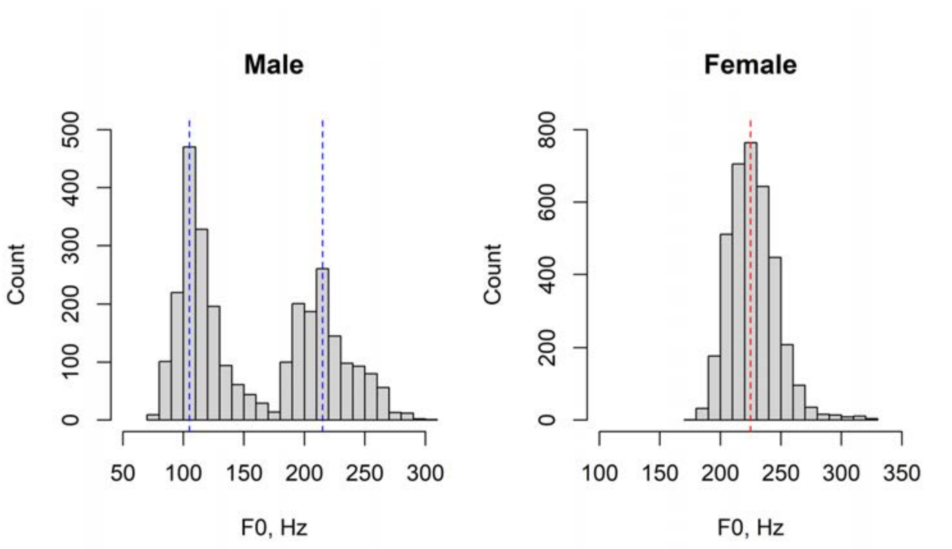

Figure 1. *Histograms of F0 mean values, boys on the left, girls on the right.*

Figures 2 and 3 represent the boxplots of the measured F0 means for each age group in boys and girls, respectively. In boys, the general developmental pattern of F0-mean values shows a gradual decrease between the ages 9 and 12 by *ca* 17 Hz, a more prominent decline (by *ca* 28 Hz) occurs between the ages 12–13 followed by the largest drop (by 51 Hz) between the ages 13–14, a further decrease (by 21 Hz) continues till age 15. During the ages of 15–18, the F0 mean stabilizes around 110 Hz (in pairwise comparison, the differences between the age groups are still statistically significant). However, individual F0 developmental paths can be different from a general pattern, as the outliers in Figure 2 manifest. E.g., there are two 13-years old (F0 means 101 and 113 Hz), three 14-years old (F0 means 228, 194, and 164 Hz), two 15-years old (F0 means 149 and 141 Hz), and one 16-years old (F0 mean 201 Hz) boys whose F0 mean values deviate significantly ($p<0.001$) from the other speakers in the respective age group (F0 means 199, 114, 105, and 111 Hz, respectively). We suggest that these deviating F0 mean values reveal an early (in the case of 13-year-olds) and late (in the case of 14–16-year-olds) beginning of the pubertal voice change period.

In girls, F0 development shows an almost linear decline pattern between the ages 9 and 18, with F0 mean decrease from ca 245 Hz to 212 Hz (Figure 3). The outliers occurring in several age groups can be attributed to the individual peculiarities of the subject's larynx.

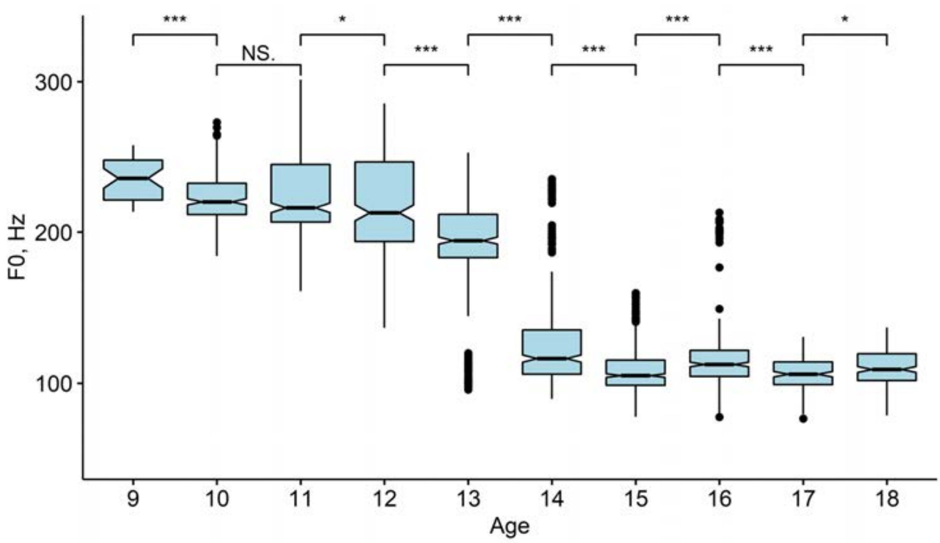

Figure 2. *Male speakers' boxplots of F0 means by age with significance levels of pairwise t-test (NS. $p > 0.05$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$).*

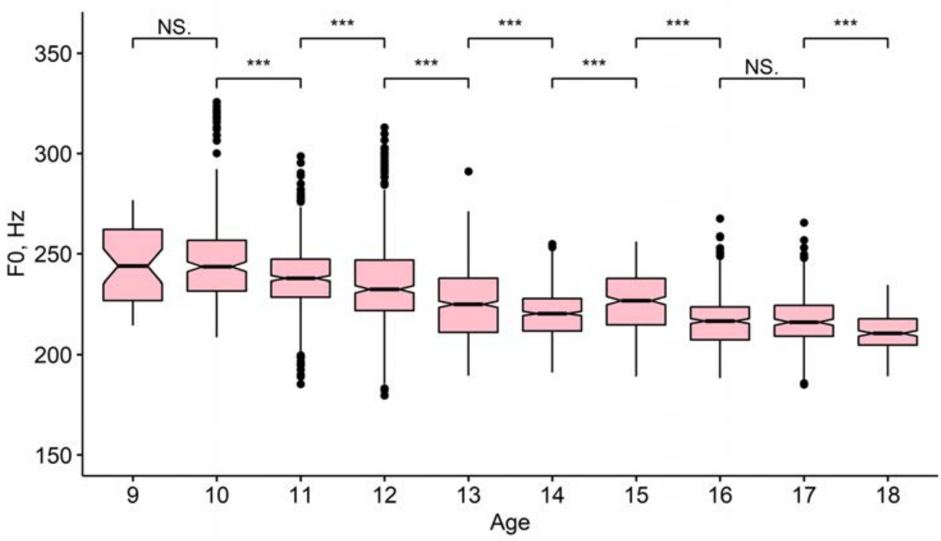

Figure 3. *Female speakers' boxplots of F0 means by age with significance levels of pairwise t-test (NS.* $p > 0.05$, * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$*).*

The measured F0 mean values for each read utterance were used to fit the GAMMs for both gender groups with the smooth term *age* and *subject* as the random effect. The model-predicted mean F0 values and the standard errors for each age group by gender are presented in Table 4 and in Figure 4.

Table 4. *The GAMM-predicted mean F0 values and the standard errors (SE) for each age group by gender (in Hz).*

		9	10	11	12	13	14	15	16	17	18
Female	F0	250	244	238	233	228	224	221	218	215	212
	SE	4.4	2.7	1.6	1.2	1.1	1.1	1.3	1.6	2.2	3.4
Male	F0	230	225	224	216	184	134	111	112	110	108
	SE	15.6	6.2	4.9	5.5	4.8	4.8	4.4	5.1	6.7	7.8

According to the GAMM's prediction, in girls between the ages of 9 and 18 years, the F0 mean gradually decreases from 250 Hz to 212 Hz. In boys between the ages 9–12, the F0 mean decreases from 230 Hz to 216 Hz, followed by a rapid drop of F0 (by 105 Hz) between the ages 12–15, with the most significant decline (by 51 Hz) occurring between 13 and 14 years; the minor differences of F0 at the ages of 15–18 show the natural subject-specific variations.

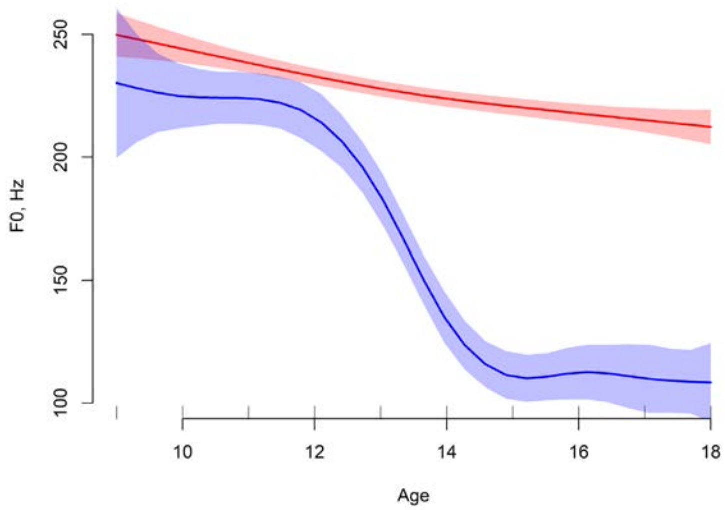

Figure 4. *The age-dependent change of F0 mean with 95% confidence bands for girls (red) and boys (blue) as predicted by the GAMMs.*

Formant frequencies of vowels

GAMMs were fitted for the vowel formants F1 and F2 with the smooth terms *age* depending on the factors *gender, vowel, quantity,* and *vowel position*, and *subject* as the random effect. The models for F1 and F2 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 5, GAMMs' numeric summaries are given in Table 5.

In both F1 and F2, the largest changes in formant frequencies occur between the ages of 10 and 15, with significant differences between boys and girls ($p<0.001$). In ages 15–18, the changes in formant values are small, and the vowels acquire a stable quality. For all vowels, the formant frequencies of girls are higher than the corresponding values of boys in the same age group, and gender differences become significant at 12–13 years of age ($p<0.05$). There is a significant effect of *quantity* ($p<0.001$) and *vowel position* ($p<0.001$) on both vowel formants revealing that overlong and long vowels are more peripheral than short ones, and the quality of vowels in unstressed syllables tend to be reduced compared to the counterparts in the stressed syllables. Thus, in adolescent speech, the factors *quantity* and *vowel position* have similar effects on vowel quality as reported for adult speech (Eek & Meister 1998; Lippus et al. 2013).

Table 5. *Estimated parametric coefficients for the factor variables and approximate significance of the smooth terms from GAMMs for F1 and F2. The reference levels for the factors are: Gender – Male, Vowel – /a/, Quantity – short, Vowel position – V1.*

Formant	Parametric coefficients				
F1	Factor	Estimate	Std. Error	t-value	p-value
	(Intercept)	5.97	0.0256	232.71	<0.001
	Vowel /e/	-0.64	0.0089	-71.72	<0.001
	Vowel /i/	-2.19	0.0073	-299.55	<0.001
	Vowel /o/	-0.96	0.0105	-91.61	<0.001
	Vowel /u/	-1.84	0.0083	-221.77	<0.001
	Vowel /õ/	-1.30	0.0171	-76.27	<0.001
	Vowel /ä/	0.417	0.0193	21.68	<0.001
	Vowel /ö/	-1.20	0.0232	-51.99	<0.001
	Vowel /ü/	-2.14	0.0137	-155.96	<0.001
	Quantity long	-0.12	0.0122	-10.12	<0.001
	Quantity overlong	-0.207	0.0106	-19.45	<0.001
	Gender Female	0.542	0.0333	16.28	<0.001
	Vowel V2	0.169	0.0060	28.07	<0.001
	Approximate significance of smooth terms:				
		edf	Ref.df	F	p-value
	s(Age)	3.804	3.826	54.27	<0.001
	s(Subject)	291.944	302	39.76	<0.001
Formant	Parametric coefficients				
F2	Factor	Estimate	Std. Error	t-value	p-value
	(Intercept)	10.39	0.0268	388.19	<0.001
	Vowel /e/	1.94	0.0143	135.79	<0.001
	Vowel /i/	3.26	0.0144	227.36	<0.001
	Vowel /o/	-1.30	0.0148	-87.89	<0.001
	Vowel /u/	-1.55	0.0125	-123.20	<0.001
	Vowel /õ/	0.32	0.0295	10.78	<0.001
	Vowel /ä/	1.52	0.0278	54.73	<0.001
	Vowel /ö/	2.14	0.0452	47.46	<0.001
	Vowel /ü/	2.12	0.0316	67.18	<0.001
	Quantity long	-0.41	0.0198	-20.58	<0.001
	Quantity overlong	-0.62	0.0181	-34.23	<0.001
	Gender Female	0.69	0.0340	20.15	<0.001
	Vowel V2	0.17	0.0103	16.60	<0.001
	Approximate significance of smooth terms:				
		edf	Ref.df	F	p-value
	s(Age)	3.667	3.727	69.18	<0.001
	s(Subject	278.958	302	13.67	<0.001

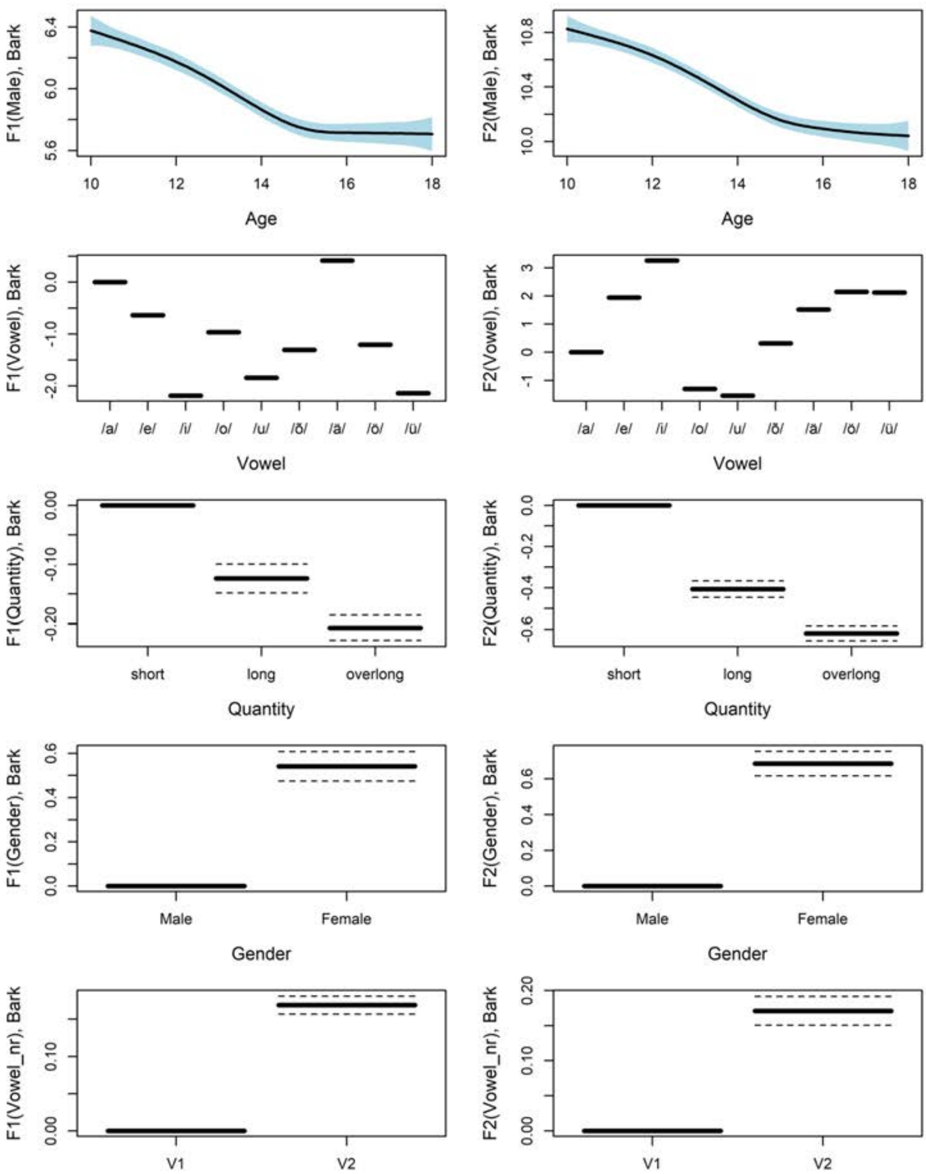

Figure 5. *The term plots of the GAMMs for F1 (left column) and F2 (right column) of adolescent vowels. The upper row represents the fitted F1 and F2 values depending on Age (shaded area corresponds to ±1 standard error), further rows represent F1 and F2 variations depending on the factors Vowel (the reference level is /a/), Quantity (the reference level is short), Gender (the reference level is Male), and Vowel_nr (the reference level is V1), respectively.*

Vowel trajectories

The numeric values of developmental vowel trajectory lengths (DVTL) of stressed vowels are presented in Table 6. Both boys and girls have the largest DVTL values for vowels /i/ and /e/ and the smallest for /o/ and /u/. Except for /u/, boys have higher DVTL values than the respective values in girls. Vowel trajectories in Figure 6 illustrate the developmental changes in the quality of stressed vowels depending on age and gender.

***Table 6**. DVTL values (in Hz and Bark) of stressed (V1) vowels for male and female speakers.*

	/a/	/e/	/i/	/o/	/u/	/õ/	/ä/	/ö/	/ü/
Male, Hz	378	592	744	182	94	245	504	418	470
Female, Hz	155	347	366	73	146	169	220	297	273
Male, Bark	1.96	2.21	2.33	1.16	0.77	1.28	2.42	1.72	1.82
Female, Bark	0.88	1.10	0.96	0.51	0.97	0.84	0.98	1.13	1.05

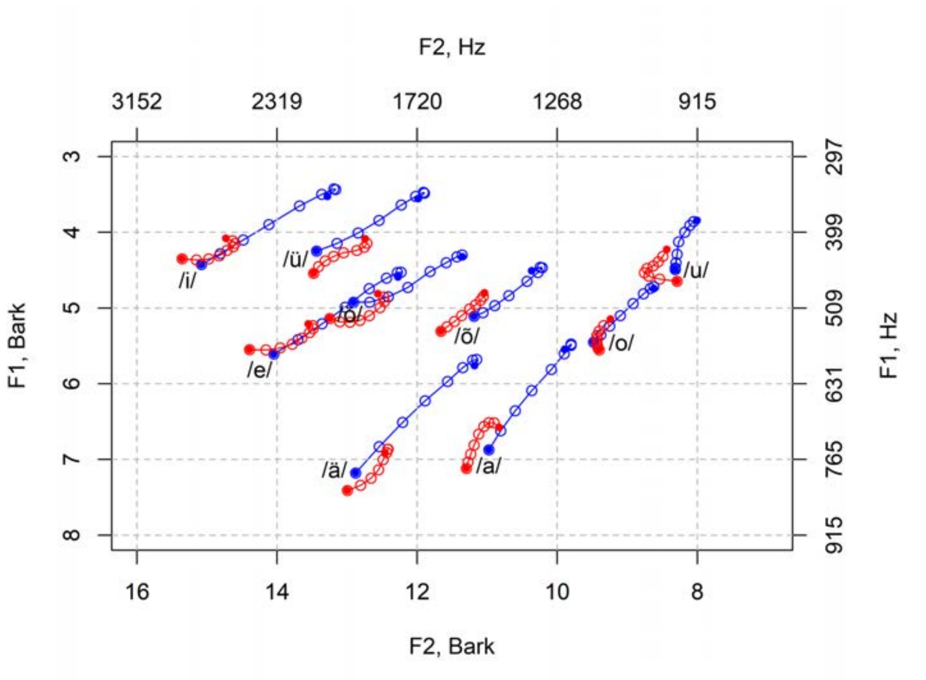

***Figure 6.** Vowel trajectories representing the age-related changes of the stressed vowels in boys (blue) and girls (red) in F1&F2 acoustic space. The filled circles represent the positions of the vowels at the age of 10 and 18 years, the empty circles represent the intermediate ages; the vowel characters are placed close to the points corresponding to the vowels of the 10-year-olds.*

For visual comparison, Figure 7 illustrates the acoustic vowel space as the area between the corner vowels of V1 (top) and V2 (bottom) vowels for 10-years-old (left column) and 18-years-old (right column) speakers.

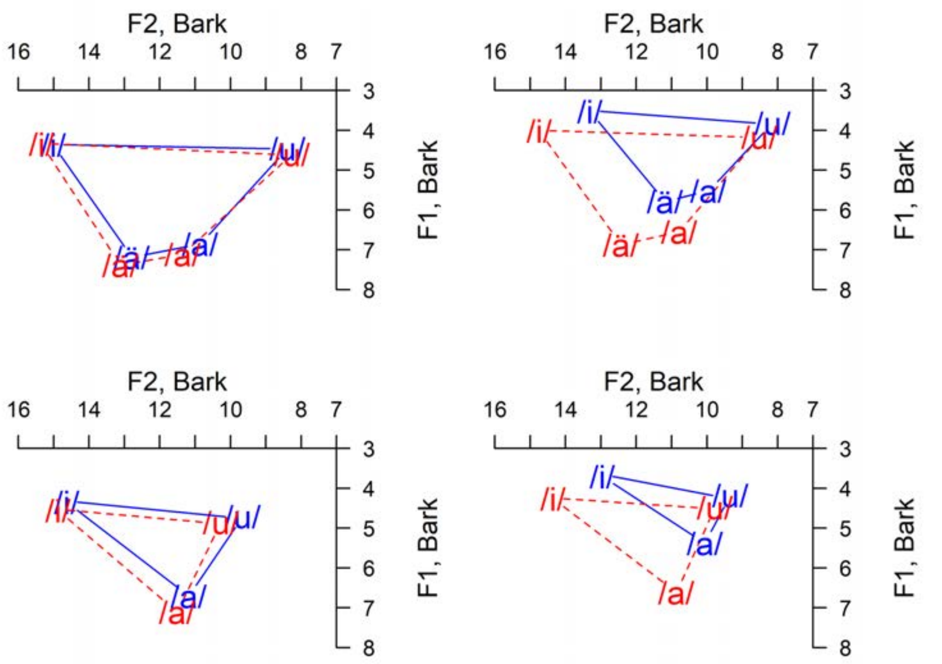

Figure 7. *The acoustic vowel space of V1 (top) and V2 (bottom) as the area between the corner vowels in boys (solid blue line) and girls (dashed read line) for 10-years-old (left column) and 18-years-old (right column) speakers.*

Vowel duration

The measured duration of V1 and V2 vowels were allocated to GAMM-modeling with the smooth terms *age,* the explanatory factors *gender*, *vowel*, *quantity* and *vowel position*; *subject* was added as a random variable. Figure 10 shows the plots of the partial effects of the explanatory features on vowel durations, the numeric summary of the GAMM is given in table 7.

Table 7. *Estimated parametric coefficients for the factor variables and approximate significance of the smooth terms from the GAMM for vowel duration. The reference levels for the factors are: Vowel – /a/, Quantity – short, Gender – Male, Vowel_nr– V1.*

Parametric coefficients				
Factor	Estimate	Std. Error	t-value	p-value
(Intercept)	83.91	0.939	89.35	<0.001
Vowel /e/	-5.20	0.378	-13.76	<0.001
Vowel /i/	-8.88	0.345	-25.75	<0.001
Vowel /o/	-0.96	0.507	-1.89	0.058
Vowel /u/	-7.23	0.401	-18.05	<0.001
Vowel /õ/	-5.75	0.825	-6.97	<0.001
Vowel /ä/	-4.28	0.716	-5.98	<0.001
Vowel /ö/	2.64	2.001	1.32	0.187
Vowel /ü/	-10.24	0.737	-13.90	<0.001
Quantity long	65.91	0.994	66.30	<0.001
Quantity overlong	87.07	1.027	84.82	<0.001
Gender Female	-0.06	1.209	-0.05	0.962
Vowel V2	3.54	0.280	12.62	<0.001
Approximate significance of smooth terms:				
	edf	Ref.df	F	p-value
s(Age)	3.41	3.45	39.47	<0.001
s(Subject)	286.06	302	19.07	<0.001

The results show that age has a significant effect on vowel duration (p<0.001) and there are no differences between male and female speakers (p=0.962). The duration of vowels decreases from age 9 to age 14, and the variations in further ages are marginal. As expected, the factor *quantity* (with levels of short, long, and overlong) has a significant effect (p<0.001) on V1 vowel duration. Although the duration of vowels decreases with age, the long/short and overlong/short duration ratios of V1 vowels (1.8–1.9 and 2.0–2.2, respectively) stay rather stable among all age groups and are close to those reported in several studies on Estonian adult speech (1.9 and 2.5, respectively) (cf. Meister 2011: 29, Table 4). However, the adolescent overlong/short duration ratio (2.0–2.2) tends to be smaller than that of adult speech (2.5).

The duration of V2 is longer than the duration of V1 short vowels (p<0.001) and has a developmental pattern analogous to that of V1 (shortening till age 14, marginal variations thereafter). GAMM output shows significant duration differences between the reference vowel /a/ and most other vowels (p<0.001), except the vowels /o/ (p=0.058) and /ö/ (p=0.962). Duration differences reveal the intrinsic microprosodic variations – low vowels tend to be longer than high vowels (Meister & Werner 2006, 2009). In current data, this tendency is only partly evident (see the vowel panel in Figure 8), as in read speech higher prosodic levels (word, utterance) might override the microprosodic features.

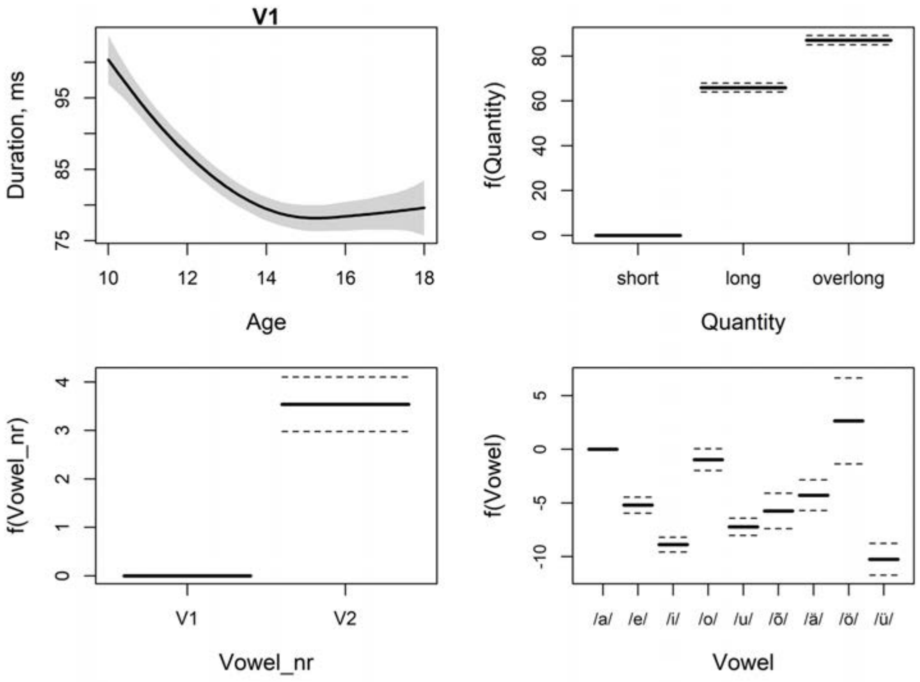

Figure 8. *The response plots of the GAMMs for vowel duration. The top row represents the model-predicted V1 duration depending on age (shaded area corresponds to ±1 standard error) and the effect of the factor Quantity (reference level short), the bottom row represents the effects of the factors Vowel_nr (reference level /V1/) and Vowel (reference level /a/).*

Speaking rate characteristics

GAMMs were fitted for the articulation and speech rates with the smooth terms *age* and *text length*, and the explanatory factor *gender*; *subject* was added to the models as an independent random variable. The numeric output of the GAMMs is given in Table 8 and the partial effects of *age* and *text length* are shown in Figure 9. The results reveal that both articulation rate and speech rate depend on subject's age ($p<0.001$) and text length ($p<0.001$), and there is no gender difference in both rates ($p=0.587$ for articulation rate and $p=0.767$ for speech rate). The difference between articulation rate and speech rate is rather small as pauses and hesitations are rare in read speech. The predicted values of the articulation and speech rates are given in Table 9.

Table 8. *Estimated parametric coefficients for the factor variables and approximate significance of the smooth terms from the GAMM for speaking rate characteristics. The reference level for the factor gender is Male.*

Articulation rate	Parametric coefficients				
	Factor	Estimate	Std. Error	t-value	p-value
	(Intercept)	5.01	0.041	122.44	<0.001
	Gender Female	-0.03	0.054	-0.543	0.587
	Approximate significance of smooth terms:				
		edf	Ref.df	F	p-value
	s(Age)	3.47	3.52	40.03	<0.001
	s(Syl_count)	4.68	4.94	273.37	<0.001
	s(Subject)	283.05	306	14.10	<0.001
Speech rate	Parametric coefficients				
	Factor	Estimate	Std. Error	t-value	p-value
	(Intercept)	4.90	0.042	115.4	<0.001
	Gender Female	-0.02	0.057	-0.3	0.767
	Approximate significance of smooth terms:				
		edf	Ref.df	F	p-value
	s(Age)	3.59	3.65	41.35	<0.001
	s(Syl_count)	4.70	4.95	188.28	<0.001
	s(Subject)	286.06	302	19.07	<0.001

Table 9. *The GAMM-predicted articulation and speech rates and the standard errors for each age group (in syllables per second).*

	9	10	11	12	13	14	15	16	17	18
Art. rate	4.0	4.3	4.6	4.9	5.0	5.2	5.3	5.3	5.2	5.2
SE	0.13	0.06	0.04	0.04	0.04	0.04	0.04	0.05	0.06	0.09
Speech rate	3.8	4.2	4.5	4.8	5.0	5.1	5.2	5.2	5.2	5.1
SE	0.14	0.07	0.05	0.04	0.04	0.04	0.04	0.05	0.06	0.09

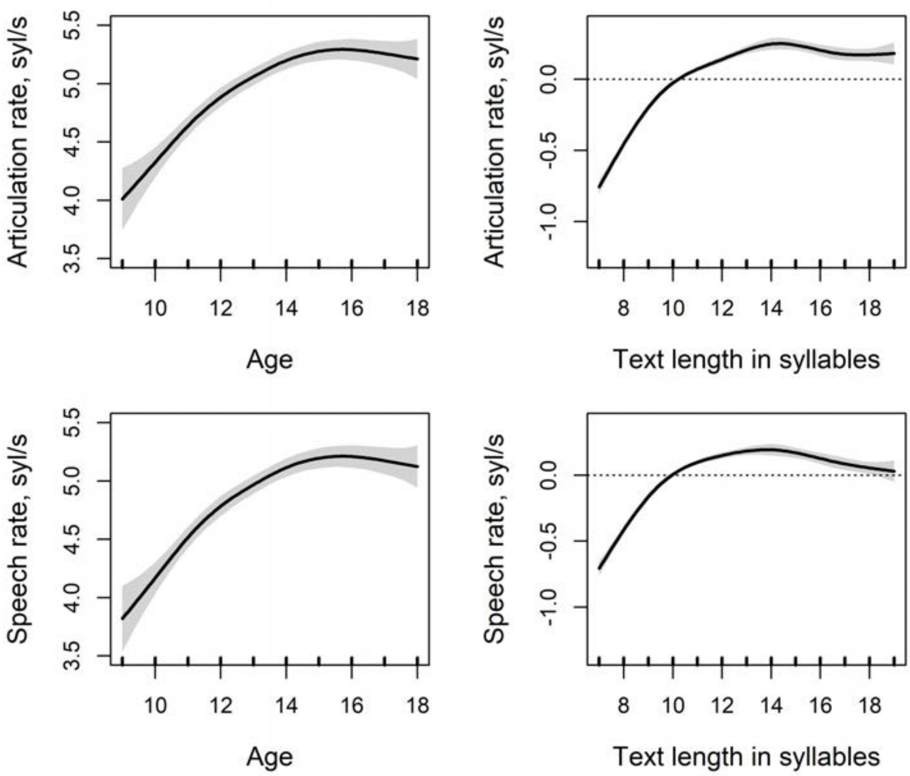

Figure 9. *The partial effects of Age and Text length fitted with GAMM models for articulation rate (top) and speech rate (bottom), shaded gray areas represent the 95% confidence interval of the mean.*

The data in Table 9 and Figure 9, left show similar age-related development patterns of both speaking rate characteristics: at ages 9–15 both rates increase significantly (pairwise comparison of age groups is statistically significant, $p<0.001$) and reach a maximum at 15 years of age, in ages 17–18 both rates slightly decrease ($p<0.05$). The effect of text length on the speaking rates has similar patterns (Figure 9, right): in sentences with 8–12 syllables the speaking rates increase with the text length (differences are significant at least at the $p<0.05$ level), in sentences with 12–18 syllables the speaking rates are stable, followed by a decreasing trend (differences are insignificant) in longer sentences (19–20 syllables).

DISCUSSION

In this study, we explored the acoustic characteristics of F0, vowel formants and duration, and speaking rate in Estonian adolescent speech. Due to the cross-sectional nature of the corpus, the findings reveal general patterns of the age- and gender-related changes along each acoustic dimension characterizing speech development.

In boys, the largest drop of F0 (by 105 Hz) occurs between age 12 (F0 = 216 Hz) and age 15 (F0 = 111 Hz). This finding is in line with Lee et al. (1999) which suggests that pubertal voice change in male speakers starts between ages 12 and 13, and ends around age 15. Similar to previous studies (Hollien et al. 1994; Whiteside et al. 2002), individual differences in 13–16-year-old males have been found in our corpus, as well, with mean F0 values that significantly differ from those of age-matched peers, suggesting that the onset of puberty varies among speakers. More detailed information on individual F0 development is provided by longitudinal studies in which participants are recorded at regular time intervals over several years (e.g. Whiteside et al. 2002; Bennet 1983; Hollien et al. 1994).

We have worked with speech material that was read in a neutral speech style with a relatively monotonous F0, therefore, the results do not represent the entire vocal range of the speakers. In different speech styles, e.g. in a spontaneous or emotional speech, higher F0 values would be expected for joy and lower for anger, the F0 range would be larger for anger and smaller for sadness as it has been found in adult speech (Tamuri 2015).

Although there are several minor differences in F0 characteristics in comparison of Estonian subjects with English- (Lee et al. 1999) and German-speaking (Draxler et al. 2008) subjects, Estonian adolescents follow similar development patterns. Studies exploring F0 differences between languages (see Patterson

2000 for a review) have found that the differences can be due to organic (e.g. differences in the vocal tract influenced by height and racial origin), linguistic (e.g. the role of F0 in the language's prosodic system), and socio-cultural factors. These factors should be considered when exploring the F0 differences between Estonian adolescents and their German- and English-speaking peers.

The variations of formant frequencies of the vowels reported in the study reflect the age- and gender-specific changes in the vocal tract of children – as age increases, the length of the vocal tract increases and, as a result, the formant frequencies of all vowels and the area of the acoustic vocal space decrease. The age-related development patterns of formant frequencies are similar in both genders, i.e. the main changes in formant frequencies occur between ages 10 and 15 and further changes are minor. Since the length of the vocal tract in boys is bigger than that of girls, the formant frequency values for boys are always lower than those of girls. Similar results have been reported in several studies on the vowel acoustics of English-speaking subjects (e.g. Flipsen & Lee 2012; Lee et al. 1999). A comparison of the lengths of the developmental vowel trajectories (Table 6) shows that the front vowels /i/ and /e/ undergo the largest quality change, while the smallest changes are in the quality of the back vowels /u/ and /o/. The reasons for these differences may lie partly in the different growth rates of different vocal tract regions, i.e. oral and pharyngeal cavities (Vorperian et al. 2009, 2011), and in the different roles of these regions in the articulation of front and back vowels. The relationships between the anatomical development of the vocal tract and its acoustic properties are more complex and non-linear, however, the changes in vowel quality cannot be fully explained by the developmental changes of the vocal tract only. In addition, e.g. the spoken language, the dialectal background of the subject (Fox & Jacewicz 2009; Jacewicz et al. 2011), and sociolinguistic factors (Pettinato et al. 2016) influence the development of vowel quality.

Unlike the variations in F0 and the spectral characteristics of vowels, changes in speech tempo are primarily related to the development of speech-motor skills. The effect of age on articulation and speech rates is apparent: both increase between the ages of 10 and 15 and become stable at further ages. According to previous studies, the temporal characteristics of children's speech are acoustically more variable up to the age of 12 years (Lee et al. 1999) and articulatorily up to the age of 14 years (Smith & Zelaznik 2004). These results have been interpreted as evidence of the achievement of adult-like speech-motor skills at 12–14 years of age (e.g., Redford & Oh 2017). On the other hand, it has been argued that the development of motor patterns continues into late adolescence (Smith 2006). In line with the latter, it has been reported that the speech rate of 13–14-year-old British children has not yet reached adult levels (Hazan & Pettinato 2014) and the speech rate of 13–17-year-old Hebrew-speaking children continues to increase (Amir & Grinfeld 2011). The average speech rate of 14–18-year-old Estonian speakers (5.1 syllables per second) compared well to the read speech rate of Estonian young adults of 4.9–5.3 syllables per second (Meister & Meister 2022). Thus, we suggest that the speech-motor control of Estonian adolescents achieves adult-like levels between the ages of 14 and 15 with further improvement of proficiency in ages 15–18.

When reading aloud written texts, speakers make significantly fewer pauses compared to spontaneous speech, therefore the differences in speech and articulation rates in the analyzed speech material are small. The difference between articulation and speech rate becomes larger when reading longer sentences as readers group the text into smaller fragments, depending on the structure of the sentence or the need to breathe, and therefore the number and length of pauses increase. The dependence of speech tempo on text length has been confirmed in many studies and is known as anticipatory shortening, according to which the speaker adjusts his average syllable duration when planning his speech according to the expected phrase length (see Bishop & Kim 2018 and references therein). In this study, the effect of text length was found to be significant for speaking rates in sentences up to 12 syllables long; when reading texts longer than these, the speech rate shows a slightly declining trend, while the articulation rate stays stable (or has a slightly increasing trend) and drops again when the sentence is longer than 18 syllables. Such patterns of speaking rate variations may be related to motor planning in text reading.

SUMMARY

The study explored the changes in F0, vowel formants and duration, and speech tempo in Estonian adolescent speech as a function of age and gender. The discovered developmental patterns are as follows: (1) a decline in F0 with a sharp drop of about 100 Hz in boys aged 12–15 years due to puberty voice change and a gradual decline in girls during ages 9–18, (2) formant frequencies of vowels decrease gradually from 10 to 15 years in both genders and the quality of vowels stabilizes at the age of 15–18 years, gender-specific differences emerge at the age of 12–13, (3) vowel duration decreases and speech tempo increases up to 15 years of age and becomes stable in further ages, gender differences are not significant. The results are in line with the findings reported for several other languages.

The results of the study will further the knowledge about the age- and gender-related variability of the acoustic properties of adolescent speech, and can be considered as reference data that are typical for Estonian-speaking individuals aged 9–18 years with normal language development.

ACKNOWLEDGEMENTS

The study was supported by the European Union through the European Regional Development Foundation (Centre of Excellence in Estonian Studies, TK 145) and by the national program Estonian Language Technology 2018–2027 (the project Speech recognition).

REFERENCES

Amir, Ofer & Grinfeld, Doreen 2011. Articulation Rate in Childhood and Adolescence: Hebrew Speakers. *Language and Speech*, Vol. 54, No. 2, pp. 225–240. https://doi.org/10.1177/0023830910397496.

Asu, Eva Liina & Lippus, Pärtel & Pajusalu, Karl & Teras, Pire 2016. *Eesti keele hääldus*. [Estonian Pronunciation.] (Eesti keele varamu II.) Tartu: Tartu Ülikooli Kirjastus.

Barbier, Guillaume & Perrier, Pascal & Payan, Yohan & Tiede, Mark K. & Gerber, Silvain & Perkell, Joseph S. & Ménard, Lucie 2020. What Anticipatory Coarticulation in Children Tells Us about Speech Motor Control Maturity. *PLoS ONE*, Vol. 15, No. 4: e0231484. https://doi.org/10.1371/journal.pone.0231484.

Bennett, Suzanne 1983. A 3-year Longitudinal Study of School-aged Children's Fundamental Frequencies. *Journal of Speech, Language, and Hearing Research*, Vol. 26, pp. 137–42. https://doi.org/10.1044/jshr.2601.137.

Bishop, Jason & Kim, Boram 2018. Anticipatory Shortening: Articulation Rate, Phrase Length, and Lookahead in Speech Production. *Proceedings of 9th International Conference on Speech Prosody*, pp. 235–239. https://doi.org/10.21437/SpeechProsody.2018-48.

Boersma, Paul & Weenink, David 2022. *Praat: doing phonetics by computer.* (Computer Program.) Version 6.3. Available at http://www.praat.org/, last accessed on 15 November 2022.

Crystal, Thomas H. & House, Arthur S. 1990. Articulation Rate and the Duration of Syllables and Stress Groups in Connected Speech. *Journal of the Acoustical Society of America*, Vol. 88, No. 1, pp. 101–112. https://doi.org/10.1121/1.399955.

Darling-White, Meghan & Banks, Symone W. 2021. Speech Rate Varies with Sentence Length in Typically Developing Children. *Journal of Speech, Language, and Hearing Research*, Vol. 64, No. 6S, pp. 2385–2391. https://doi.org/10.1044/2020_JSLHR-20-00276.

Draxler, Christoph & Jänsch, Klaus 2004. SpeechRecorder – a Universal Platform Independent Multi-Channel Audio Recording Software. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04), Lisbon*, pp. 559–562.

Draxler, Christoph & Schiel, Florian & Ellbogen, Tania 2008. F0 of Adolescent Speakers – First Results for the German Ph@ttSessionz Database. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. European Language Resources Association. Available at http://www.lrec-conf.org/proceedings/lrec2008/, last accessed on 3 February 2023.

Eek, Arvo & Meister, Einar 1998. Quality of Standard Estonian Vowels in Stressed and Unstressed Syllables if the Feet in Three Distinctive Quantity Degrees. *Linguistica Uralica*, Vol. 34, No. 3, pp. 226–233.

Fant, Gunnar 1960. *Acoustic Theory of Speech Production*. The Hague, Netherlands: Mouton.

Fitch, Tecumseh W. & Giedd, Jay 1999. 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. https://doi.org/10.1121/1.427148.

Flipsen, Peter jr. & Lee, Sungbok 2012. Reference Data for the American English Acoustic Vowel Space. *Clinical Linguistics & Phonetics*, Vol. 26, No. 11–12, pp. 926–933. https://doi.org/10.3109/02699206.2012.720634.

Fox, Robert A. & Jacewicz, Ewa 2009. Cross-Dialectal Variation in Formant Dynamics of American English Vowels. *Journal of the Acoustical Society of America,* Vol. 126, No. 5, pp. 2603–2618. https://doi.org/10.1121/1.3212921.

Goffman, Lisa & Smith, Anne 1999. Development and Differentiation of Speech Movement Patterns. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 25, No. 3, pp. 649–660. https://doi.org/10.1037/0096-1523.25.3.649.

Green, Jordan R. & Moore, Christopher A. & Higashikawa, Masahiko & Steeve, Roger W. 2000. The Physiologic Development of Speech Motor Control: Lip and Jaw Coordination. *Journal of Speech, Language, and Hearing Research*, Vol. 43, No. 1, pp. 239–255. https://doi.org/10.1044/jslhr.4301.239.

Hazan, Valerie & Pettinato, Michèle 2014. The Emergence of Rhythmic Strategies for Clarifying Speech: Variation of Syllable Rate and Pausing in Adults, Children and Teenagers. *10th International Speech Production Seminar, Köln, Germany*, pp. 178–181.

Hirano, Minoru & Kurita, Shigejiro & Toh, Yuichi 1981. Growth, Development and Aging of the Vocal Fold. *Practica Oto-Rhino-Laryngologica*, Vol. 74, No. 8, pp. 1791–180. https://doi.org/10.5631/jibirin.74.1791.

Hirst, Daniel 2007. A Praat Plugin for Momel and INTSINT With Improved Algorithms for Modelling and Coding Intonation. *Proceedings of the XVIth International Congress of Phonetic Sciences*, pp. 1233–1236.

Hollien, Harry & Green, Rachel & Massey, Karen 1994. Longitudinal Research on Adolescent Voice Change in Males. *Journal of the Acoustical Society of America*, Vol. 96, No. 5, pp. 2646–2654. https://doi.org/10.1121/1.411275.

Ingham, Janis Costello & Riley, Glyndon 1998. Guidelines for Documentation of Treatment Efficacy for Young Children Who Stutter. *Journal of Speech, Language, and Hearing Research*, Vol. 41, No. 4, pp. 753–770. https://doi.org/10.1044/jslhr.4104.753.

Jacewicz, Ewa & Fox, Robert A. & O'Neill, Caitlin & Salmons, Joseph 2009. Articulation Rate Across Dialect, Age, and Gender. *Language Variation and Change*, Vol. 21, No. 2, pp. 233–256. https://doi.org/10.1017/S0954394509990093.

Jacewicz, Ewa & Fox, Robert A. & Salmons, Joseph 2011. Vowel Change Across Three Age Groups of Speakers in Three Regional Varieties of American English. *Journal of Phonetics*, Vol. 39, No. 4, pp. 683–693. https://doi.org/10.1016/j.wocn.2011.07.003.

Jacewicz, Ewa & Fox, Robert A. & Wei, Lai 2010. Between-Speaker And Within-Speaker Variation In Speech Tempo of American English. *Journal of the Acoustical Society of America*, Vol. 128, No. 2, pp. 839–50. https://doi.org/10.1121/1.3459842.

Lee, Alice & Doherty, Rachel 2017. Speaking Rate and Articulation Rate of Native Speakers of Irish English. *Speech, Language and Hearing*, Vol. 20, No. 4, pp. 206–211. https://doi.org/10.1080/2050571X.2017.1290337.

Lee, Sungbok & Potamianos, Alexandros & Narayanan, Shrikanth 1999. Acoustics of Children's Speech: Developmental Changes of Temporal and Spectral Parameters. *Journal of the Acoustical Society of America*, Vol. 105, No. 3, pp. 1455–1468. https://doi.org/10.1121/1.426686.

Lehiste, Ilse 1960. Segmental and Syllabic Quantity in Estonian. *American Studies of Uralic linguistics*. Bloomington: Indiana University, pp. 21–82.

Lehiste, Ilse 1974. Interaction Between Test Word Duration and Length of Utterance. *Working Papers in Linguistics*, No. 17, pp. 160–169.

Lippus, Pärtel 2015. *plugin_PhonCorpTools.* (Computer program.) Available at https://gitlab.keeleressursid.ee/partel/plugin_PhonCorpTools/-/blob/master/lisa/SKK_silbitamine.praatscript, last accessed on 12 June 2020.

Lippus, Pärtel & Asu, Eva Liina & Teras, Pire & Tuisk, Tuuli 2013. Quantity-related Variation of Duration, Pitch and Vowel Quality in Spontaneous Estonian. *Journal of Phonetics*, Vol. 41, No. 1, pp. 17–28. https://doi.org/10.1016/j.wocn.2012.09.005.

Logan, Kenneth J. & Byrd, Courtney T. & Mazzocchi, Elizabeth M. & Gillam, Ronald B. 2011. Speaking Rate Characteristics of Elementary-school-aged Children Who Do and do Not Stutter. *Journal of Communication Disorders*, Vol. 44, No. 1, pp. 130–147. https://doi.org/10.1016/j.jcomdis.2010.08.001.

McCloy, Daniel R. 2016. *phonR: tools for phoneticians and phonologists*. (Computer program.) R package version 1.0-7.

Meister, Einar & Meister, Lya 2022. Eesti laste kõne III: kõnetempo ja silbikestuste analüüs. [Estonian Adolescent Speech III: An Analysis of Speech Tempo and Syllable Durations.] *Keel ja Kirjandus*, No. 3, pp. 226–244. https://doi.org/10.54013/kk771a3.

Meister, Einar & Werner, Stefan 2006. Intrinsic Microprosodic Variations in Estonian and Finnish: Acoustic Analysis. *The Phonetics Symposium 2006, Department of Speech Sciences, University of Helsinki*, Vol. 53, pp. 103–112.

Meister, Einar & Werner, Stefan 2009. Duration Affects Vowel Perception in Estonian and Finnish. *Linguistica Uralica*, Vol. 45, No. 3, pp. 161–177. https://doi.org/10.3176/lu.2009.3.01.

Meister, Lya 2011. *Eesti vokaali- ja kestuskategooriad vene emakeelega keelejuhtide tajus ja häälduses. Eksperimentaalfoneetiline uurimus*. [The Perception and Production of Estonian Vowel and Duration-Based Categories by Non-Native Subjects With a Russian-Language Background. An Experimental Phonetic Study.] Tartu: University of Tartu.
Meister, Lya & Meister, Einar 2014. Development of the Corpus of Estonian Adolescent Speech. *Human Language Technologies – the Baltic Perspective. Proceedings of the Sixth International Baltic Conference, Baltic HLT 2014*. (Frontiers in Artificial Intelligence and Applications, 268.) Amsterdam: IOS Press, pp. 206–209. https://doi.org/10.3233/978-1-61499-442-8-206.
Patterson, David J. 2000. *A Linguistic Approach to Pitch Range Modelling*. PhD. thesis, Edinburgh: University of Edinburgh.
Pettinato, Michèle & Tuomainen, Outi & Granlund, Sonia & Hazan, Valerie 2016. Vowel Space Area in Later Childhood and Adolescence: Effects of Age, Sex and Ease of Communication. *Journal of Phonetics*, Vol. 54, pp. 1–14. https://doi.org/10.1016/j.wocn.2015.07.002.
R Core Team 2018. *R: A Language and Environment for Statistical Computing*. R. Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Redford, Melissa A. & Oh, Grace E. 2017. The Representation and Execution of Articulatory Timing in First and Second Language Acquisition. *Journal of Phonetics*, Vol. 63, No. 1, pp. 127–138. https://doi.org/10.1016/j.wocn.2017.01.004.
Robb, Michael P. & Maclagan, Margaret A. & Chen, Yang 2004. Speaking Rates of American and New Zealand Varieties of English. *Clinical Linguistics & Phonetics*, Vol. 18, No. 1, pp. 1–15. https://doi.org/10.1080/0269920031000105336.
Rogers, Derek J. & Setlur, Jennifer & Raol, Nikhila & Maurer, Rie & Hartnick, Christopher J. 2014. Evaluation of True Vocal Fold Growth as a Function of Age. *Otolaryngology–Head Neck Surgery*, Vol. 151, No. 4, pp. 681–687. https://doi.org/10.1177/0194599814547489.
RStudio Team 2020. *RStudio: Integrated Development for R*. Boston, MA: RStudio, PBC. Available at http://www.rstudio.com, last accessed on 20 June 2020.
Sadagopan, Neeraja & Smith, Anne 2008. Developmental Changes in the Effects of Utterance Length and Complexity on Speech Movement Variability. *Journal of Speech, Language, and Hearing Research*, Vol. 51, No. 5, pp. 1138–1151. https://doi.org/10.1044/1092-4388(2008/06-0222).
Schötz, Susanne & Frid, Johan & Löfqvist, Anders 2013. Development of Speech Motor Control: Lip Movement Variability. *Journal of Acoustical Society of America*, Vol. 133, No. 6, pp. 4210–4217. https://doi.org/10.1121/1.4802649.
Sharkey, Susan G. & Folkins, John W. 1985. Variability of Lip and Jaw Movements in Children and Adults: Implications for the Development of Speech Motor Control. *Journal of Speech and Hearing Research*, Vol. 28, No. 1, pp. 8–15. https://doi.org/10.1044/jshr.2801.08.
Smith, Anne 2006. Speech Motor Development: Integrating Muscles, Movements, and Linguistic Units. *Journal of Communication Disorders*, Vol. 39, No. 5, pp. 331–349. https://doi.org/10.1016/j.jcomdis.2006.06.017.

Smith, Anne & Goffman, Lisa 1998. Stability and Patterning of Speech Movement Sequences in Children and Adults. *Journal of Speech, Language, and Hearing Research*, Vol. 41, No. 1, pp. 18–30. https://doi.org/10.1044/jslhr.4101.18.

Smith, Anne & Zelaznik, Howard N. 2004. Development of Functional Synergies for Speech Motor Coordination in Childhood and Adolescence. *Developmental Psychobiology*, Vol. 45, No. 1, pp. 22–33. https://doi.org/10.1002/dev.20009.

Stevens, Kenneth N. 2000. *Acoustic Phonetics.* Cambridge, Massachusetts, London, England: The MIT Press.

Tamuri, Kairi 2015. Fundamental Frequency in Estonian Emotional Read-Out Speech. *Journal of Estonian and Finno-Ugric Linguistics*, Vol. 6, No. 1, pp. 9–21. https://doi.org/10.12697/jeful.2015.6.1.01.

Tsao,Ying-Chiao & Weismer, Gary & Iqbal, Kamran 2006. Interspeaker Variation in Habitual Speaking Rate: Additional Evidence. *Journal of Speech, Language, and Hearing Research*, Vol. 49, No. 5, pp. 1156–1164. https://doi.org/10.1044/1092-4388(2006/083).

Van Rij, Jacolien & Wieling, Martijn & Baayen, Harald & van Rijn, Hedderik 2022. *"itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs."* R package version 2.4.1. Available at https://cran.r-project.org/web/packages/itsadug/, last accessed on 6 February 2023.

Verhoeven, Jo & De Pauw, Guy & Kloots, Hanne 2004. Speech Rate in a Pluricentric Language: A Comparison Between Dutch in Belgium and the Netherlands. *Language and Speech*, Vol. 47, No. 3, pp. 297–308. https://doi.org/10.1177/00238309040470030401.

Vorperian, Houri K. & Kent, Ray D. 2007. 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. https://doi.org/10.1044/1092-4388(2007/104).

Vorperian, Houri K. & Wang, Shubing & Chung, Moo K. & Schimek, Michael E. & Durtschi, Reid B. & Kent, Ray D. & Ziegert, Andrew J. & Gentry Lindell R. 2009. Anatomic Development of the Oral and Pharyngeal Portions of the Vocal Tract: An Imaging Study. *Journal of the Acoustical Society of America*, Vol. 125, No. 3, pp. 1666–1678. https://doi.org/10.1121/1.3075589.

Vorperian, Houri K. & Wang, Shubing & Schimek, Michael E. & Durtschi, Reid B. & Kent, Ray D. & Gentry, Lindell R. & Chung, Moo K. 2011. 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. https://doi.org/10.1044/1092-4388(2010/10-0097).

Whiteside, Sandra P. & Hodgson, Carolyn & Tapster, Caroline 2002. Vocal Characteristics in Pre-adolescent and Adolescent Children: A Longitudinal Study. *Logopedics Phoniatrics Vocology*, Vol. 27, No. 1, pp.12–20. https://doi.org/10.1080/140154302760146934.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R.* Second Edition. CRC Press. https://doi.org/10.1201/9781315370279.

Einar Meister (PhD) is senior research fellow at the Laboratory of Language Technology, Department of Software Science, Tallinn University of Technology. His research areas are Estonian phonetics, L2 speech, speech corpora, speech analysis and synthesis, and speech technology applications. In recent years, he has focused on sociophonetic studies, including age- and gender-related variations in adolescent and elderly speech, and singing voice.

einar.meister@ttu.ee

Lya Meister (PhD) is research fellow at the Laboratory of Language Technology, Department of Software Science, Tallinn University of Technology. Her research includes phonetic studies on Estonian native and L2 speech, age- and gender-related variations in adolescent and elderly speech, and most recently singing voice. She has collected several speech corpora for phonetic studies and speech recognition.

lya.meister@ttu.ee