Human Language Technologies – The Baltic Perspective
A. Utka et al. (Eds.)

doi:10.3233/978-1-61499-442-8-206

Development of the Corpus of Estonian Adolescent Speech

Lya MEISTER [a,1], Einar MEISTER [a]
[a] *Institute of Cybernetics at Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia*

Abstract. The paper presents the work in progress on the development of the Corpus of Estonian Adolescent Speech. The corpus comprises speech recordings of native subjects between 9 and 18 years of age recorded in several public schools in different dialect regions of Estonia. It will include samples from 300 subjects balanced by gender and age. Preliminary analysis of F0 provides new information on voice development of adolescent subjects.

Keywords. Adolescent speech, Estonian, fundamental frequency

Introduction

Since the 1980s, the creation of large speech corpora has become increasingly important for phonetic analysis and even more important for training and testing speech technology systems. Especially, in speech recognition the approach – *there is no data like more data* – is still valid in the current state-of-the art technology. The existing Estonian speech resources [1], [2] available for the training of speech recognition systems include the recordings of adult subjects only. Therefore, child and adolescent speech remains challenging for the existing speech recognition systems. In addition, there exist no suitable corpora for socio-phonetic studies addressing the development of voice and speech of this age group. These were the two main motives for the development of the Corpus of Estonian Adolescent Speech that was started in 2011 under the national program Estonian Language Technology (2011-2017)".

The corpora of adolescent speech collected for German [3] and Scottish [4] have served as prototypes for the current project.

1. Corpus Specification

1.1. Corpus Design

The corpus includes speech recordings of native Estonian subjects in the age range from 9 to 18 years and it aims at: (1) covering the age group that is not at all represented in the existing Estonian speech corpora, (2) serving as a resource for speech technology

[1] Corresponding Author: Lya Meister; E-mail: lya@phon.ioc.ee.

(mainly for training of speech recognition systems) as well as for sociolinguistic and phonetic studies. Ultimately, the corpus will include samples from 300 subjects with balanced gender and age distribution, about 20 minutes of speech from each subject.

1.2. Corpus Content

The corpus contains linguistically diverse material: digits, numbers, phone numbers, time and date expressions, IT terms, sentences with place, person and institution names, phonetically rich sentences, two longer passages (see Table 1). Samples of spontaneous speech are elicited with standard prompts including self-introduction and topic suggestions for storytelling (about school, hobbies, etc.), and pictures to be described. The diversity of the corpus content ensures the compatibility with the existing Estonian adult speech corpora and makes it possible to compare adolescent and adult speech using coherent and partly verbatim speech material. Several phonetically rich sentences of the current corpus are derived from the Estonian Babel Corpus [1]. They involve all Estonian vowels and consonants as well as frequent diphthongs and consonant clusters in two-syllable word structures representing Estonian quantity contrasts. Application-oriented (time and date expressions, digits, PIN-codes, telephone numbers, IT terms) are compatible with the Estonian SpeechDat Corpus [2].

Table 1. Corpus description.

Item description	**Count**	
	per session	total
Standard items	10	10
Time and date expressions	5	90
Digits, numbers, PIN-codes	9	270
Phone numbers	4	150
IT terms	5	150
Phonetically rich sentences	21	90
Sentences with place names	3	200
Sentences with person names	3	200
Sentences with object names	5	150
Passages	2	60
Pictures	3	15
Total	**70**	**1,385**

1.3. Recording Procedure

The recordings are carried out using a mobile recording set including a laptop with BAS SpeechRecorder software [5], a microphone preamplifier with USB interface (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 are stored directly to the hard disc in wav format (sampling at 44.1 kHz, resolution 16 bit). At the beginning of each recording session the recording process is explained to a subject and some test prompts are recorded in order to adjust the signal levels from both microphones. Recording session is supervised by a technical operator, and when necessary, additional explanations are given to the subject during the session.

2. Data Collection

For the recruitment of subjects several schools in the capital area and in three dialectal areas of Estonia (North-East and South Estonia, and Saaremaa) were approached. The schools were selected on the basis of their willingness to cooperate with the recording team and of the availability of a suitable room for recordings (usually a quiet classroom). The selection of subjects was made by the schools, typically by a teacher of Estonian according to given criteria (native Estonian, balance by age and gender, no hearing and speaking disorders, fluency in reading of unfamiliar texts). Approvals from head teacher and from parents were obtained beforehand. From each subject the following data was collected: date of birth, gender, school and grade, mother tongue, other languages studied, place of living in early childhood, and current place of living.

In general, the quality of most recordings is good. However, as a post-recording quality check of randomly selected signals revealed, in some recorded signals several dropped frames at irregular intervals, occasional clipping, and rather high level of background noise were found. The clipping has occurred due to variable reading style of subjects, usually greater variability of loudness occurring in spontaneous speech. In school environment the background noise is picked up by the microphones especially during lesson breaks when pupils run and communicate loudly in corridors, also the room acoustics (e.g. signal reflection from the walls) causes additional degradation of signal quality.

Since all subjects had never participated in such recordings, they were rather exited about the opportunity to participate in the project. However, especially younger pupils were often slightly stressed and worried about their performance in reading unfamiliar texts and too shy in tasks involving story telling or description of pictures. These situations needed personal approach and encouragement from the recording team in order to help the subject relax and quickly adapt to the recording situation.

Currently, recordings of 230 subjects (59% female, 41% male) have been made in four high schools in the capital area and in two schools in North-East Estonia. The recordings in Saaremaa and in South Estonia will be carried out in autumn 2014.

3. Preliminary Acoustic Analysis

Research on the development of human vocal tract has found anatomic gender differences in the oral and pharyngeal areas of the vocal tract in prepubertal, pubertal and postpubertal age groups, see e.g. [6]. As the acoustic implication of these anatomical changes, age and gender related differences in fundamental frequency (F0) and in vowel formants have been documented in different languages ([7] and references therein).

Using the recordings so far collected, we have carried out a preliminary analysis of F0 distribution depending on age and gender. From each subject a subset of read items (21 sentences) was used. A Praat [8] script was developed to calculate F0 mean, median, minimum, maximum, and standard deviation. F0 statistics were calculated first for each utterance, and then pooled over utterances to obtain the results for each subject.

For male speakers, F0 mean decreases gradually from 230 Hz to 186 Hz in the age from 9 to 12 years, due to puberty voice mutation it drops down ca 50 Hz in the age 12–13, and then it lowers further from 136 Hz to 110 Hz at the age from 13 to 18 years. For female speakers, F0 mean shows a gradual change from 250 Hz (9 years) to 210 Hz

(18 years) without such remarkable drop-down that was observed in male subjects. Also the other F0 characteristics show similar patterns. The standard deviation of F0 means shows the largest values in males of age 13 and 14 – the subjects of this age groups are probably at the end of the voice mutation period and their F0 is still rather variable. F0 range is rather stable in females at all ages; in males it narrows after the age of 14.

Similar F0 trends have been reported for German adolescents [9]. However, the voice mutation period is observed earlier in Estonian males, occurring as it does in the age interval from 12 to 14 years, whereas in the case of German males it takes place between 13 and 15 years of age.

4. Summary

We have introduced the work in progress on the development of the Corpus of Estonian Adolescent Speech and presented preliminary results on F0 of male and female subjects depending on age. Our further work will include data collection in South Estonia and Saaremaa later this year, and segmentation and labelling of the whole corpus.

The corpus will be available via the Center of Estonian Language Resources (http://keeleressursid.ee/) and via the EU CLARIN infrastructure (http://www.clarin.eu/).

5. Acknowledgements

We thank the schools that made the data collection possible and all volunteer speakers. The work has been supported by the National Program for Estonian Language Technology and by the target-financed theme No. 0140007s12 of the Estonian Ministry of Education and Research.

References

[1] A. Eek and E. Meister, Estonian speech in the BABEL multi-language database: Phonetic-phonological problems revealed in the text corpus. In: *Proceedings of LP'98*, Fujimura, O.(Ed.), Prague: The Karolinum Press, 1999, 529–546.

[2] E. Meister, J. Lasn, and L. Meister, SpeechDat-like Estonian database. In: *Text, Speech and Dialogue: 6th International Conference, TSD 2003*, Matoušek, V. and Mautner, P. (Eds.), Berlin: Springer, Lecture Notes in Artificial Intelligence **2807**, 2003, 412–417.

[3] Chr. Draxler, K. Jänsch, Speech recordings in public schools in Germany – the perfect show case for web-based recordings and annotation. In: *Proceedings of LREC 2006*, Genova, 2006.

[4] C. Dickie, F. Schaeffler, Chr. Draxler, and K. Jänsch, Speech recordings via the internet: An overview of the VOYS project in Scotland. In: *Proceedings of Interspeech2009*, Brighton, 2009.

[5] SpeechRecorder [Computer program]. Version 2.8.4, retrieved 2 May 2014 from http://www.bas.uni-muenchen.de/Bas/software/speechrecorder/.

[6] H.K. Vorperian, S. Wang, E.M. Schimek, R.B. Durtschi, 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* **54**, 2011, 995–1010.

[7] S.A. Xue, R.W, Cheng, and L.M. Ng, Vocal tract dimensional development of adolescents: An acoustic reflection study. *International Journal Of Pediatric Otorhinolaryngology*, **74**(8), 2010, 907–912.

[8] P. Boersma and D. Weenink, Praat: doing phonetics by computer [Computer program]. Version 5.3.70, retrieved 2 April 2014 from http://www.praat.org/.

[9] Chr. Draxler, F. Schiel, and T. Ellbogen, F0 Of Adolescent Speakers – First Results for the German Ph@ttSessionz Database. In: *Proceedings of LREC 2008*, Marrakech, 2008.