Age Dependence of Children's Speech Parameters

J. Janda

This paper deals with the search for agedependent parameters in children's speech. These parameters are compared in terms of age dependence, and their adequacy for recognizing the age of a speaker is presented, using discrimination analysis.

Keywords: children's speech, age dependence, phonetic analysis.

1 Introduction

An analysis of the relationship between acoustic-phonetic aspects of speech and the speaker's age may have numerous applications. This paper has been motivated by practical experience in the field of phoniatry and logopaedia. When examining children's pathological speech, there is often an effort to answer the question "What age does particular speech corresponding to", and therefore for example to estimate at what age a child's speech development stopped.

Chronological age is unambiguously given by date of birth. Logopaedic age is the age estimated on the basis of acoustic-phonetic aspects of human speech.

2 Materials and methods

For the purposes of this survey, a database of children's speech was recorded. It consists of the speech of 193 children aged from three to twelve years. It contains the following words (in Czech): *babička, časopis, čokoláda, dědeček, kalhoty, kniha, košile, květina, květiny, maluje, mateřídouška, motovidlo, peníze, pohádka, pokémon, popelnice, radost, rukavice, různobarevný, silnice, škola, špička, televize, ticho, trumpeta, vlak, zelenina, zmrzlina.*

2.1 Frequency of a basic glottic tone F0

An analysis was made of separate vowels in syllables /la/, /le/, /li/, /lo/, /lu/ from the words *škola, košile, zmrzlina, letadlo* and *maluje*, and then of complete vocal sections of the speech.

The analysis was made using an autocorrelation method in the *Praat v. 5.0.15* program [6] with the following parameters *time step* = 0.0, *pitch floor* = 100 Hz and *pitch ceiling* = 600 Hz. The resulting values were verified using the *Wavwsurfer v. 1.8.5* program [7], and were manually modified, if applicable. The most frequent event was incorrect detection of F_0, lower by one octave.

In order to make the frequency intervals comply better with the perception of intonation intervals by human hearing, the F_0 values were transferred to a semitone scale, with the beginning at 100 Hz

$$F_0(ST) = 12 \ln \frac{F_0(Hz)/100}{\ln(2)}. \tag{1}$$

For statistical confirmation of the age dependence of F_0, a zero hypothesis of H_0 was taken into consideration, which denies such dependence. H_0 can be rejected on the basis of the results of a t-test for the correlated measurements:

$$t = \frac{\left| \bar{x}_{F_0} - \bar{x}_{\text{age}} \right|}{s_{\bar{d}}}, \tag{2}$$

where

$$s_{\bar{d}} = \frac{\sqrt{\dfrac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}}{\sqrt{n}}. \tag{3}$$

The d_i variable in this case means the difference between F_0 and the age of speaker No. i.

In our case, H_0 can be rejected for the level of , $p < 0.001$, $n = 193$. The correlation power can be expressed using the Pearson correlation coefficient:

$$r = \frac{\text{cov}(x_{F_0}, x_{\text{age}})}{s_{x_{F_0}} s_{x_{\text{age}}}}. \tag{4}$$

For the age dependence of F_0 for vowel /a/: $r = 0.43$, which is a mesoscale satisfactory correlation. For all vocal sections of speech: $r = 0.41$ with $p < 0.001$, $n = 113$. The F_0 trend is shown in Fig. 1.
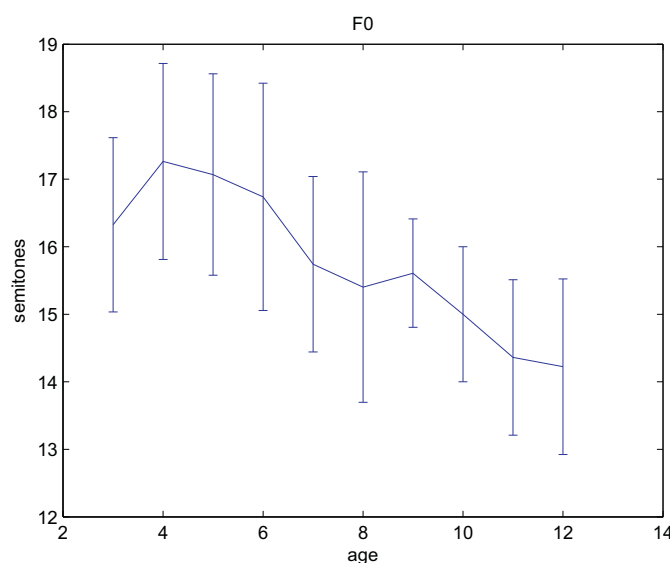

Fig. 1: F0 age dependence – complete speech

2.2 F0 variance

The variance of the basic voice frequency is associated with the intonation range of a piece of speech. This parameter reflects the overall tunefulness and melodiousness of the speech typical of pre-school children.

The F0 variance was analysed for all vocal speech sections, and showed a declining tendency with age. The correlation coefficient was $r = -0.61$ ($p < 0.001$, $n = 113$).

2.3 F1, F2 formants

Formant frequencies correspond with the resonance frequencies of the vocal organ cavities [1]. They were estimated for particular vowels using an LPC (linear predictive coding) spectrum via an algorithm by Burg [6].

Age dependence was less evident with the formants than with F0. Within all speech, F1 had a correlation coefficient of $r = -0.25$ ($p < 0.001$, $n = 193$) and F2 had $r = -0.34$ ($p < 0.001$, $n = 113$).

2.4 Sibilant consonant characteristics

Spectral centre of gravity

If the complex spectrum is given by $S(f)$, where f is the frequency, the centre of gravity is given by

$$f_c = \int_0^\infty f |S(f)|^2 \, \mathrm{d}f \qquad (5)$$

divided by the energy

$$\int_0^\infty |S(f)|^2 \, \mathrm{d}f. \qquad (6)$$

Thus, the centre of gravity is the average of f over the entire frequency domain, weighted by the power spectrum.

Central spectral moment

The n-th central spectral moment is given by

$$\mu_n = \frac{\int_0^\infty (f - f_c)^n |S(f)|^2 \, \mathrm{d}f}{\int_0^\infty |S(f)|^2 \, \mathrm{d}f}. \qquad (7)$$

Thus, the n-th central moment is the average of $(f - f_c)^n$ over the entire frequency domain, weighted by the power spectrum.

Spectral standard deviation

The standard deviation of a spectrum is the square root of the second central moment of this spectrum.

Skewness of a spectrum

The (normalized) skewness of a spectrum is the third central moment of this spectrum, divided by the 1.5 power of the second central moment.

Skewness is a measure for how greatly the shape of the spectrum below the centre of gravity differs from the shape above the mean frequency. For white noise, the skewness is zero.

Kurtosis of a spectrum

The (normalized) kurtosis of a spectrum is the fourth central moment of this spectrum, divided by the square of the second central moment, minus 3.

Kurtosis is a measure for how greatly the shape of the spectrum around the centre of gravity differs from a Gaussian shape. For white noise, the kurtosis is $-6/5$.

The above-mentioned spectral characteristics of sibilant consonants were measured for consonants /s/, /ss/ and /cc/. Significant features were especially the rise in the spectral centre of gravity ($r = 0.45$, $p < 0.001$, $n = 193$) (Fig. 3, 4) and the reduction in spectral skewness ($r = -0.47$, $p < 0.001$, $n = 193$) (Fig. 2) at consonant /s/ in word "silnice".
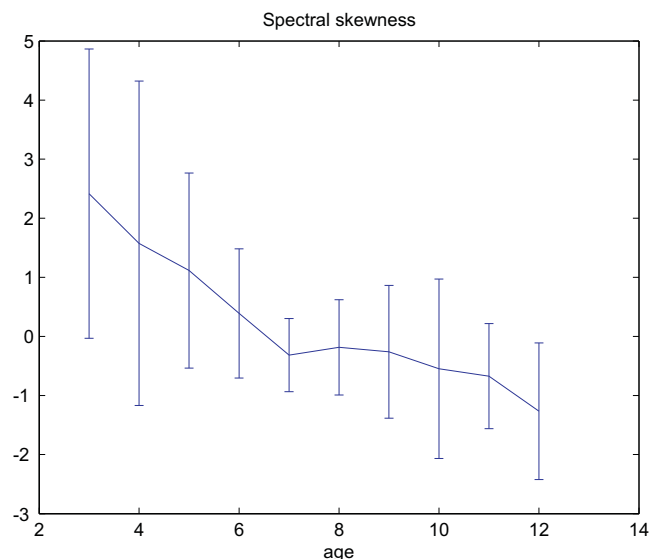
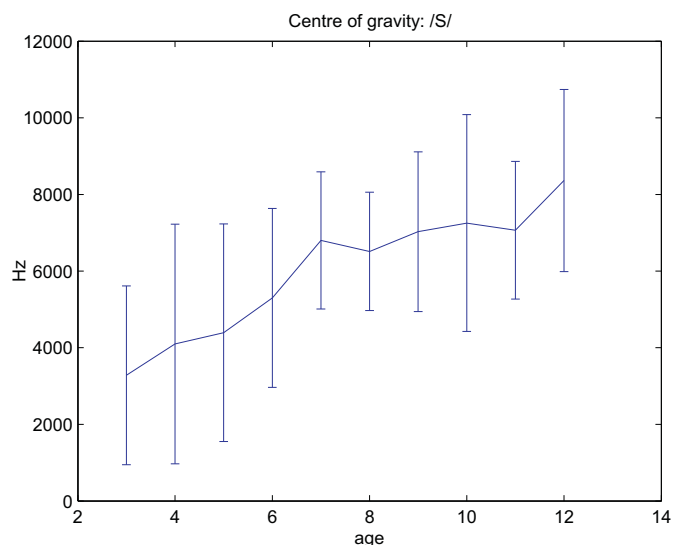Fig. 2: Age dependence of spectral skewness for consonant /s/

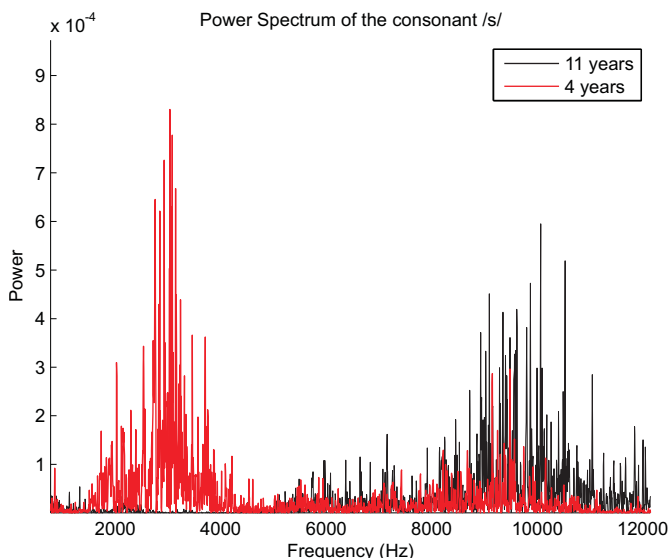Fig. 3: Spectral centre of gravity shift

Fig. 4: Spectral centre of gravity shift

2.5 Voice onset time

Voice Onset Time (VOT) [5] is the time duration between the release of a plosive and the beginning of vocal cord vibration (Fig. 5). This period is measured in milliseconds (ms).

VOT measurements were performed on syllable /ka/ from word "*babička*". However, it was not possible to prove any age dependence even of this parameter using the measured values on the level of $p < 0.05$.

2.6. Speech rate

Speech rate has been determined for particular talkers as a reciprocal value of the duration of the entire speech without pauses. Age dependence was also not proved for this parameter.

3 Results

3.1 Overview of age dependent parameters

The table below summarizes the examined phonetic characteristics. Individual attributes are ordered according to the age-correlation rate (column r). The H_0 column contains significance level values where it is theoretically possible to reject the zero hypothesis of age independent parameters. The parameters below the double line cannot be considered age-dependent on the significance level of 5 %.

Table 1: Overview of age dependent features

Feature	r	H_0
F0 variation	−0.61	9.3E−13
Spectral skewness /S/	−0.47	5.0E−12
Spec. centre of gravity /S/	0.45	8.7E−11
F0 – whole discourse	−0.42	4.0E−06
F2 – whole discourse	−0.34	1.9E−04
Spec. standard deviation /CC/	−0.30	1.4E−03
Spec. standard deviation /S/	−0.21	3.2E−03
Spec. standard deviation /SS/	−0.20	4.6E−03
Spectral kurtosis /S/	−0.17	1.8E−02
Spectral skewness /SS/	−0.14	4.9E−02
Spec. centre of gravity /SS/	0.11	1.4E−01
Spec. centre of gravity /CC/	−0.12	1.9E−01
Spectral kurtosis /CC/	0.11	2.6E−01
Spectral skewness /CC/	0.10	2.8E−01
Voice onset time /K-A/	−0.08	3.7E−01
Spectral kurtosis /SS/	0.00	9.6E−01
Speech rate	0.00	9.8E−01
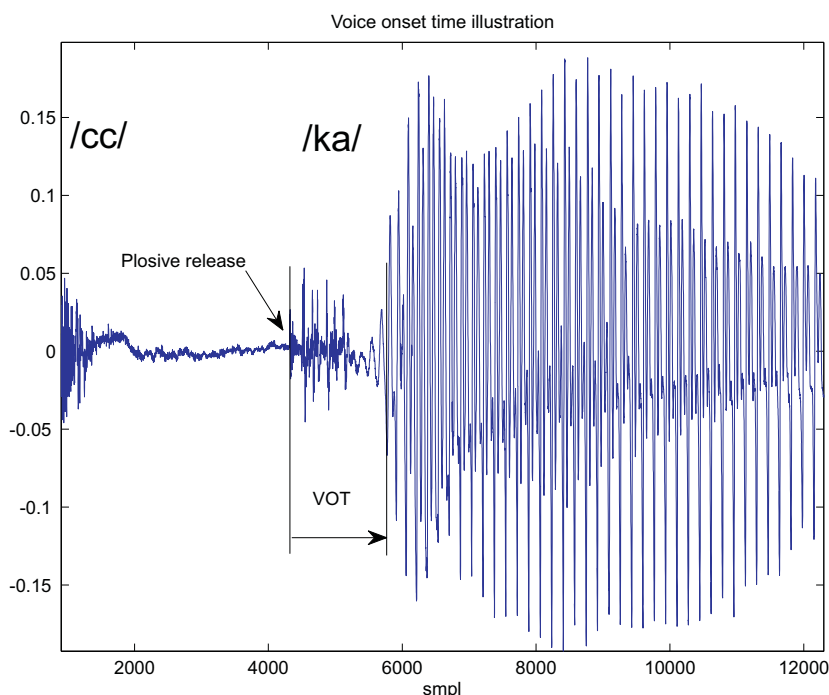
Fig. 5: Voice onset time illustration. Syllable /ka/ from word "babička"

http://ctn.cvut.cz/ap/

3.2 Discrimination analysis

In this part, we will try to make use of the age-dependent parameters for a simple discrimination analysis. The data classification is based on acceptance or rejection of the hypothesis of data pertinence to a particular class. Four classes were designated (**0**: 3–5 years, **1**: 6–7 years, **2**: 8–9 years, **3**: 10–12 years). The discrimination function being maximized is as follows [2]:

$$g_i(\mathbf{d}) = \ln p(\mathbf{d}|h_i) = -\ln|\mathbf{C_i}| - (\mathbf{d} - \mu_i)^T \mathbf{C_i}^{-1}(\mathbf{d} - \mu_i), \qquad (8)$$

where $\mathbf{C_i}$ is the covariance matrix, μ_i is the mean value and $(\mathbf{d}|h_i)$ is the probability rate of the results on **d** data, on the assumption that hypothesis h_i applies.

Training was performed using the RANSAC method for the vectors of 16 phonetic parameters.

The classification success rate is shown in Fig. 6, and the percentage enumeration is shown in a confusion matrix (Table 2).
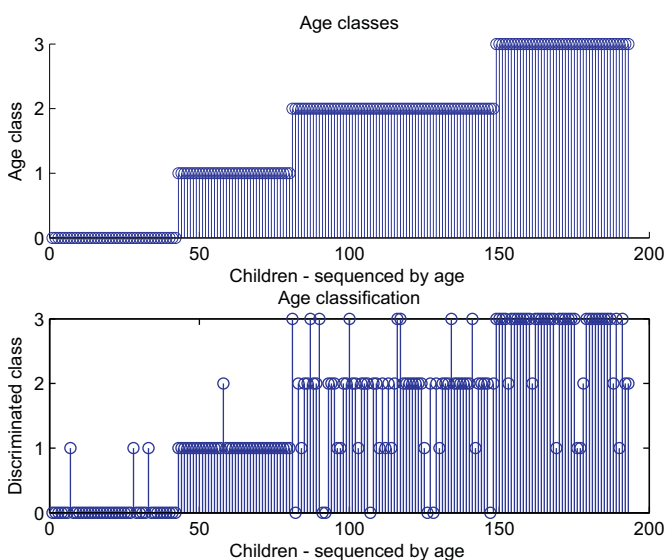

Fig. 6: Age classification

Table 2: Confusion matrix

Actual	Predicted			
	0	*1*	*2*	*3*
0	92.9	7.1	0.0	0.0
1	0.0	97.4	2.6	0.0
2	10.3	14.7	63.2	11.8
3	0.0	8.9	13.3	77.8

4 Conclusion

The selected speech characteristics showed various intensities of age dependence. The characteristics based on basic vocal frequency and some spectral properties of consonant /s/ showed a correlation of about 0.5. In the end, it was shown that selected speech attributes enable training of a classifier

which provides for classification into age groups with a probability rate of ca 80 % A similar classification method will be tested in the future on the speech og children with speech developmental defects.

References

[1] Psutka, J. et al.: *Mluvíme s počítačem česky*. Prague: Academia, 2006.

[2] Uhlíř, J., Sovka, P. et al.: *Technologie hlasových komunikací*. Prague: CTU – Publishing House, 2007, ISBN 978-80-01-03888-8

[3] Ohnesorg, K.: *Naše dítě se učí mluvit*. Prague: SPN, 1976, ISBN 80-04-25233-8.

[4] Schötz, S.: Acoustic Analysis of Adult Speaker Age. In *Speaker Classification I*. Heidelberg: Springer-Verlag, 2007.

[5] Whiteside, S. P., Marshall, J.: Developmental Trends in Voice Onset Time: Some Evidence for Sex Differences. *Phonetica*, Vol. **58**, No. 3. p. 196–210.

[6] Boersma, P., Weenink, D.: *Praat: Doing Phonetics by Computer* (Version 4.3.14), 2005, [Computer program]. http://www.praat.org/

[7] Sjölander, K., Beskow, J.: *WaveSurfer* [Computer program], http://www.speech.kth.se/wavesurfer/

Jan Janda
e-mail: jandaj2@fel.cvut.cz

Department of Circuit Theory

Czech Technical University in Prague
Faculty of Electrical Engineering
Technická 2
166 27 Prague, Czech Republic