Perceptual and acoustic analysis of voice similarities between parents and young children

Evgeniia Rykova
Université Toulouse Paul Sabatier
rykova.eugenia@gmail.com

Stefan Werner
University of Eastern Finland
stefan.werner@uef.fi

Abstract

Human voice provides the means for verbal communication and forms a part of personal identity. Due to genetic and environmental factors, a voice of a child should resemble the voice of her parent(s), but voice similarities between parents and young children are underresearched. Read-aloud speech of Finnish-speaking and Russian-speaking parent-child pairs was subject to perceptual and multi-step instrumental and statistical analysis. Finnish-speaking listeners could not discriminate family pairs auditorily in an XAB paradigm, but the Russian-speaking listeners' mean accuracy of answers reached 72.5%. On average, in both language groups family-internal f0 similarities were stronger than family-external, with parents showing greater family-internal similarities than children. Auditory similarities did not reflect acoustic similarities in a straightforward way.

1 Introduction

The current paper is based on the research made as a master thesis. An overall inspiration comes from encountering online the company VocaliD Inc., whose aim is to create unique personalized voices for text to speech devices (VocaliD, Inc.). The author asked herself, "How would a (hypothetical) voice of a child, who never had an ability to speak, most likely sound?" Intuitively, it should somehow resemble the voice of the parent(s). However, the up-to-date research does not give a direct answer to the question. In the present paper, the similarity between parents and their young children is also researched from the cross-linguistic perspective, comparing two prosodically different patterns.

2 Background

2.1 Human voice similarities

Human voice, a sound produced by a combination of human organs called vocal apparatus, is used by humans to generate speech and other forms of vocalizations. Each voice is unique due to the physiological factors (e.g., age, body size or hormones) and the manner in which the sounds are articulated (consciously or unconsciously). Due to the same factors, the voice of an individual is subject not only to major changes throughout the lifespan (Decoster and Debruyne, 2000; Stathopolous et al., 2011), but also in everyday communication. Thus, it is a source of biological, psychological and social (Bogdanova, 2001; Bolinger, 1989) information about the speaker. Both related and unrelated people can sound alike. In the blood members of the same family, the reasons for such similarities are both biological (genetic) and environmental. The former are reflected not only in the body parts but also in structural brain organization (Peper et al., 2007; Thompson et al., 2001). The latter include socialization and learning by imitation (Zuo and Mok, 2015; see also Hirvonen, 1970; Bolinger, 1989). Interestingly, the prosody of the native language is acquired earlier than the segmental phonology (Iivonen, 1977) and around two years of age, children are able to produce adult-like intonational contrasts (Bolinger, 1986).

Juslin and Scherer (2008) divide the cues for voice description into four broad groups, as related to 1) fundamental frequency (f0); 2) intensity; 3) temporal aspects; 4) voice quality. Acknowledging the importance of all the voice cues in building voice identity of an individual, for the purposes of the current research, f0 (or its contour, a sequence of f0 values across an

utterance) will be the principal feature in focus. F0 analysis is a robust acoustic method of speaker identification (Labutin et al., 2007; Rose, 1999) and the source for prosody generation in speech synthesis. Linguistically, f0 encodes suprasegmental categories of tone, stress and intonation (Rose, 1999). F0 contour is the most important physical correlate of intonation (Iivonen, 2005).

Primarily mean f0 shows significantly high intra-twin correlation in monozygotic twins, (Debruyne et al., 2002; Decoster et al., 2001; Fuchs et al., 2000; Przybyla et al., 1992; Van Lierde et al., 2005). Dizygotic twins show greater discrepancies in f0 than monozygotic twins (Debruyne et al., 2002; Przybyla et al., 1992), but the same f0 variation, which is thus considered to correspond to learnt language behavior (Debruyne et al., 2002). A variety of studies on perceptual similarity also show that twins, followed by same-sex siblings, are the most difficult to differentiate both for human listeners and an automatic system (Decoster et al, 2001; Feiser and Kleber, 2012; Kushner and Bickley, 1995; Nolan et al., 2011; Rose and Duncan, 1995; Rose, 1999; San Segundo and Kunzel, 2015; Sebastian et al., 2013; Weirich and Lancia, 2011). Listeners are also able to identify twin and sibling pairs in different tasks, and in general rate voices of related speakers with higher similarity scores than those of unrelated speakers. In most of the experiments, longer utterances seem to be more suitable stimuli. Albeit one word is enough to distinguish unrelated speakers in the study by Weirich and Lancia (2011); when the voices are knowingly similar-sounding, the task becomes more difficult even for familiar listeners (Rose and Duncan, 1995; Rose, 1999). F0 seems to be one of the most important factors that contribute to detect similarity between speakers, on one hand, and to determine dissimilarity, on the other.

2.2 Finnish and Russian prosody/intonation

A detailed comparison of phonetics, phonology and phonotactics is far beyond the scope of the current paper. In brief, Finnish is a mora-timed language with primary stress is fixed to the initial syllable of the word. Russian is stress-timed with movable word stress. Unlike Finnish, Russian has vowel reduction and no phonological durational contrasts (see, e.g., Suomi et al., 2008 and Zvukovaya forma, 2001-2002, respectively). A typical property of Finnish is falling or rising-falling intonation, steadily and smoothly declining, so that Finnish is often called prosodically monotone (Suomi et al., 2006; 2008). Russian language, on the opposite, presents a variety of f0 falling and rising contrasts with floating intonation center (Bryzgunova, 1977; Nikolaeva, 1970; Volskaya, 2009; see also Ullakonoja et al., 2007 for a comparison). Intonation in Russian plays a distinctive role in structures, where in Finnish, grammatical means are sufficient to express the difference and the difference can be characterized as mostly pragmatical (de Silva and Ullakonoja, 2009).

3 Method

3.1 Audio-data collection

The current paper presents the analysis of data collected from three mother-child pairs, whose native language is Finnish (parents of mean age 43.67 y/o, SD=4.93; two girls of 10 and 12 y/o and one nine-year-old boy), and four pairs, whose native language is Russian (parents of mean age 41.5 y/o, SD=2.65 and 12-year-old girls). The participants had no history of neurological, language or speech deficits, had normal or corrected-to-normal vision and were right-handed. They were monolingual, with some knowledge of foreign languages, but the native language being the only one spoken at home.

The young age of the boy allows to include his voice/f0 into analysis together with the girls. Mutation, or significant f0 lowering ,shows the first signs on average at the age of 10-11 (Hacki and Heitmüller, 1999). Additionally, boys before puberty might speak at a higher f0 with mothers that with fathers (Bolinger, 1989).

The recording of audio-data consisted of reading a text and five short dialogues (20 sentences of different types in total) and producing quasi-spontaneous speech in a picture description task, but only the read-aloud speech was further analyzed acoustically.

The members of the same family were recorded together. The text was first read by the child, then read by the parent in order to promote her own way of reading it and decrease the imitation effect. The dialogues were read in pairs. The recordings were made at 44100 Hz sampling frequency, and 16-bit bit depth. The

files were saved in wav-format[1] and later segmented into separate sentences. The Finnish families are coded with letters H, L and P, and the Russian families are coded with letter combinations AL, MA, OO and VN.

3.2 Perceptual experiments

Young (from 20 to 30 y/o, M=26.08, SD=2.68) native speakers of Finnish and Russian (twelve and fifteen, respectively, gender-balanced) were asked to judge the perceptual similarities in the families. They had no history of neurological, language, speech or hearing deficits, and had normal or corrected-to-normal vision.

The perceptual experiments in both languages consisted of two parts. In the first part, a participant first heard an item, pronounced by a child, followed by a beep-signal, and then the same item, pronounced by two adults, one of which was the child's parent (target) and the other served as a distractor. The task was to choose the adult, whose voice sounded more likely to be that of the child's parent. In the second part, the task was the opposite: to choose the child, whose voice sounded more likely to be that of the adult's offspring. There were training trials in each part, and the test trials (36 as 3 families x 6 items x 2 in Finnish, and 40 as 4 families x 10 items in Russian) were randomized and could be repeated three times each. Scoring was binary. The audio was presented binaurally, the experiments were conducted in a quiet environment.

3.3 Instrumental and statistical analysis

First, all the segmented sentences were compared pairwise in the same family in order to find auditory and gross f0 curve similarities. The corresponding recordings were annotated in TextGrid files. All the selected sentences in Finnish resulted to follow a falling pattern (see Iivonen, 1978; Anttila, 2009) and therefore were annotated at syllable and word level only, without distinguishing between sentence types. Annotation of the Russian data included the following: (1) section: (prepeak) – peak – (tail); (2) movement: rise/fall/rise-fall; (3) position: non-final/focus (the part containing IC of the sentence)/final; (4) group: subject or predicate; (5) orthographic word; (6) sentence type. Segmenting sentences into positions adapts the principle of additional syntagmatic segmentation (Bryzgunova, 1977). Segmenting into sections adapts the principle of a tone unit structure (see Brazil et al., 1980; Crystal, 2003): a prepeak corresponds to the pre-head and head, and a peak corresponds to the nucleus. After comparing the f0 contours inside each word for the Finnish data, and inside each position for the Russian data, the most similar pairs of sentences were chosen for the following analysis.

The f0 contours of the sentences were described through the following values: maximum f0 of the first syllables, min f0 of the other syllables for the Finnish data; maximum f0 of the peaks, mean and minimum f0 of the prepeaks and tails for the Russian data.

Since the selected Finnish sentences had different number of words and the words had different number of syllables, an equal framework of five three-syllable words was created. Thus, each word was represented by three data points, hereinafter referred to as syllables (1-3), unless otherwise specified. The syllables represent raw initially extracted values or means of the adjacent values that were close to each other. The same principle of "adjacent similarities" was applied to make five-word sentences out of six-word sentences. Missing values of the syllables were added manually following the dependencies shown between the similarly positioned syllables in the speech of the speaker. Such manipulations were applied within identical patterns in family pairs.

Statistical analysis was performed in R (R Core Team, 2017). For the purposes of the current study, analysis of variance (ANOVA) and a posthoc Tukey's Honest Significant Difference (THSD) tests were used. All the tests were carried out at 95% confidence. The graphs were created via ggplot function (Wickham, 2009).

As shown by ANOVA tests, in the Finnish data word position had a statistically significant effect (p-values (p) less than 0.05) on the raw f0 values, while the interaction word*sentence did not (p's greater than 0.1). Therefore, the words from different sentences were compared to each

1 Recording, segmentation, instrumental analysis and perceptual experiment were carried out via Praat (Boersma and Weenink, 2017).

other in accordance with their position (1-5). In the Russian data, the position*sentence interaction was similarly non-significant (p's greater than 0.1), but the effect of the position was significant (p's less than 0.05) for six out of eight speakers. The comparison of the same positions from different sentences was nevertheless applied to all the analyzed data.

The f0 features were scanned for similarities within each family (general speech rhythm comparison). However, each child within a language group was not only compared to their parent, but to all the parents in question (and vice versa) by means of ratios, calculated dividing the f0 values of each word/position from the selected sentences pronounced by an adult by the f0 values of the same words/positions from every selected sentence produced by children, data point by data point. The ratios were selected for the further analysis on the grounds of their homogeneity (0.1 as the maximum difference between the values) within a word/position and, additionally, visual similarity between f0 curves. The exception was made for some individual high peaks in the Russian data. The ratios were considered acceptable if the peak value was more than two standard deviations higher than the adjacent segments in the data from both speakers in question.

Finally, the selected ratios were reviewed word by word or position by position, focusing on the statistical differences in each pair of speakers. The ratios without significant differences were clustered together. The clusters were characterized with a coefficient, which was the mean of the clustered ratios, and strength, which was the number of clustered ratios. The latter was interpreted as the strength of similarity between the speakers. The strongest clusters from each pair of speakers were further compared to each other and used for creation of the "sentence maps", examples of which are presented in the following section.

4 Results

4.1 Perceptual experiments

In the Finnish data, none of the explanatory variables or their interactions show significant effect on the results (p's greater than 0.1 in a series of ANOVA tests). In the parent-matching task, mean accuracy per target ranges from 50% to 61.8%, M=56%, SD=5.9%; and the accuracy of answers per participant ranges from 44.4% to 63.9%, M=56%, SD=6.5%. In the child-matching task, mean accuracy per target ranges from 51.4% to 57.6%, M=53.5%, SD=3.6%; and the accuracy of answers per participant ranges from 33% to 72%, M=53.5%, SD=1.7%.

In the Russian data, the accuracy of answers per participant ranges from 50% to 77.5%, M=65%, SD=8.9% in the parent-matching task; from 50% to 80%, M=66.8%, SD=7% in the child matching task. The ANOVA tests show a significant effect of the target on the answer accuracy in both tasks (p's less than 0.01). In the parent-matching task, mean accuracy of answers for target AL (42%) is significantly lower than for the other targets (range from 68% to 78.7%, M=72%, SD=5.8%). In the child-matching task, mean accuracy of answers for target MA (49.3%) is significantly lower than for the other targets (range from 71.3% to 76.7%, M=73%, SD=2.9%). In the child-matching task, there is also a significant effect of the distractor: mean accuracy of answers with VN distractor is higher than with AL-distractor, adjusted p=0.04. Item and item*target interaction have statistically significant effect in both parts of the experiment: F=3.392, p=0.005, Df=5 and F=9.972, p =2.63e-14, Df=4, respectively, in the parent-matching-task; F=5.082, p=4.96e-04, Df=4 and F=7.448, p=2.44e-10, Df=9, respectively, in the child-matching task. THSD test shows that for every target the distribution of mean accuracy among the items is different. In other words, the same item corresponds to different mean accuracy for different targets.

The effect of language on the results of the perceptual experiment is obvious (F=26.73, p=2.57e-07, Df=1 in the ANOVA test). Task (F value=0.074; p=0.785) and interaction task*language (F=1.549; p=0.213, Df=1) do not show a significant effect on the results

4.2 Family-internal f0 similarities

For each Finnish speaker, ANOVA test shows a significant effect of the syllable and word, but not of their interaction on the f0. The adjacent similarities between the words and syllables inside the words are based on the difference between mean f0 values.

Figure 1 presents the similarities graphically: if the difference between mean f0 values is not statistically significant (adjusted p greater than 0.05 in a THSD test), the adjacent syllables/words are united with a circle. The adjusted p's at the edge of significance are marked with symbols.

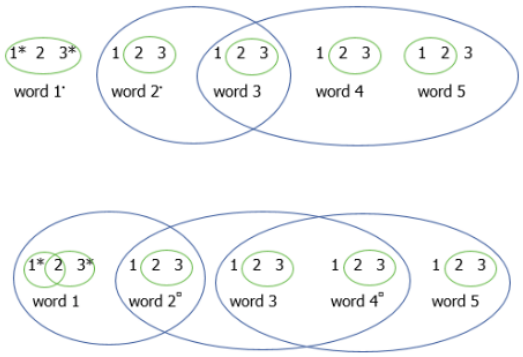

Figure 1. H-child's (above) and H-parent's (below) adjacent syllable and word f0 similarities.

Certain syllable groupings seem to appear mostly as a feature of the language, not showing great differences among all the speakers, while the word groupings seem to be more characteristic of a speaker. The absolute values of the mean f0 differences between syllables do not seem to differ that much from each other; however, the statistical significance of the difference between syllable 2 and syllable 3 varies for every speaker. The strongest child-parent similarity is found in L-family, while in families H and P parents' adjusted p-values are at the edge of significance in comparison to the children's. Majorly, the strongest adjacent similarities in children and their parents resemble each other, while the differences are found in the weakest ones. However, the child-parent dissimilarities manifest themselves differently among the families.

For each Russian speaker, the ANOVA tests do not show a significant effect of the move, so all the curve shapes inside each position are analyzed together. There is a significant effect of the section and position on the f0 values. For some speakers, the ANOVA tests also show effect of the group, but a series of THSD tests reveal that the underlying difference is between the positions. Similarly to the Finnish data, the similarities between the positions and sections inside the positions are based on the difference between mean f0 values. Figure 2 presents the similarities graphically.

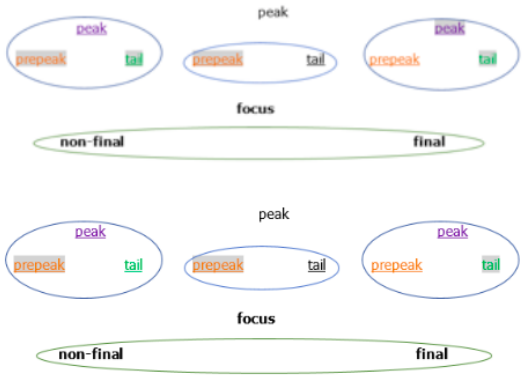

Figure 2. OO-child's (above) and OO-parent's (below) section and position f0 similarities.

Three out of four families show similar patterns of gross similarities among the positions, and each family has its own similarities and differences inside positions and cross-positionally. All of the significant differences have the same direction (sign) in both speakers of each pair.

4.3 Cluster analysis and sentence maps

In the Finnish data, the ANOVA test shows a significant effect of relationship between an adult and a child on the strength of clustered ratios: F=47.15, p=4.03e-11, Df=1. According to a THSD test, the mean strength of the same family-internal clusters is greater than that of the different pairs by 1.43. In fact, family-internal similarities on average are either stronger (H-family members, P-child) or non-significantly different (L-family members, P-parent) in comparison to the respective member's external similarities. The similarities with alien family members can be weaker (importantly, it holds absolutely true for L-parent) or non-significantly different in comparison to the latter's family-internal similarities. In a word-wise comparison, the similarities in families are in total stronger than the similarities of their members with the others in 55%, non-significantly different in 30% and weaker in 15% of the cases.

In the Russian data, The ANOVA test also shows a significant effect of the relationship between an adult and a child on the strength of clustered ratios: F=149.6; p<2e-16, Df=1. According to a THSD test, mean strength of the

family-internal clusters is greater than that of the unrelated speakers by 3.47. Family-internal similarities on average are either stronger (AL-family members, VN-family members, OO-parent) or non-significantly different (MA-family members, OO-child) in comparison to the respective member's external similarities. The similarities with alien family members can be weaker (importantly, it holds absolutely true for MA-family members) or non-significantly different in comparison to the latter's family-internal similarities. In a position-wise comparison, the similarities in families in total are stronger than the family-external similarities of their members in 81%, non-significantly different in 8% and weaker in 11% of the cases.

Besides the strength of the clusters (similarities), their coefficients and the homogeneity of the latter through a sentence are an important factor of the parent-child resemblance for both Finnish and Russian speakers. Figure 3 displays the sentence map of H-parent – H-child (HH) clusters.

For HH speaker combination, maximum possible word grouping is five words, clusters [1B + 2A* + 3D + 4C + 5C] with the mean syllable-wise coefficients [0.755; 0.746; 0.745]. The difference of 0.03 between the means of 1B and 2A, however, is at the edge of significance, adjusted p=0.03; while in the rest of the pair-wise comparisons adjusted p's are greater than 0.1. The total strength of the grouped clusters, or the sum of the maximum cluster strengths from each element, equals 28.

Figure 4 displays the sentence map of AL-parent - AL-child (ALAL) clusters. For ALAL speaker combination, maximum possible position grouping is clusters [non-final C, focus A, final A] with the mean section-wise coefficients [0.716; 0.725; 0.733; 0.723; 0.713]. The total strength of the grouped clusters equals 27.

Relation (F=157.17, p-value<2e-16, Df=1; F=144.44, p-value <2e-26, Df=1), language (F=31.49, p=2.95e-08, Df=1; F=6.915, p=0.01), and relation*language interaction (F=33.19, p=1.28e-08, Df=1; F=4.235, p=0.042, Df=1) have a significant effect both on the strength of the clusters (adult-child similarity) and total strength of groupings (statistic values given respectively). For both measures, the strength is greater in pairs of the same family members in general, family-internally greater in Russian, and family-externally greater in Finnish. The (relative) number of grouped elements is also higher in pairs of the same family members in total (adjusted p=0.002), but family-internally is higher in Finnish, and family-externally higher in Russian. It is important to note that a grouping in the Russian data does not correspond to a sentence in the same sense that a grouping in the Finnish data. In Russian, neither the number of data points in positions (three or five), nor the number and the order of the latter are fixed. In Finnish, on the opposite, the framework used in the current study reflects the permanent number and the order of the words and syllables.

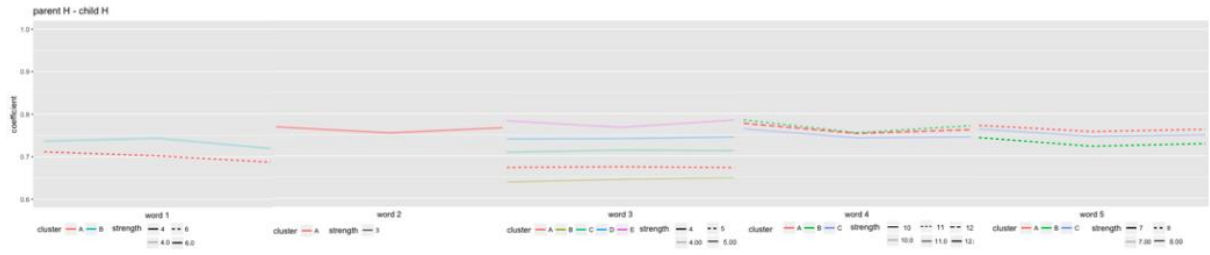

Figure 3. Sentence map of HH-clusters.

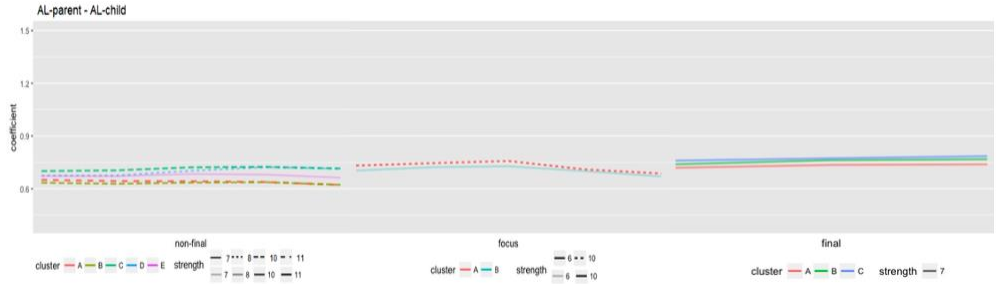

Figure 4. Sentence map of ALAL clusters.

4.4 Relationship between the perceptual experiments and instrumental analysis

ANOVA tests, run with the binary result of the perceptual test as a dependent variable, and the total strength of groupings divided into two explanatory variables, similarity (family-internal) and distracting power (family-external), show the only significant effect of the similarity on the accuracy results in the child-matching task in Russian (F=13.3, p=3.00e-04, Df=1), which reflects that significantly lower accuracy for target MA is associated with its low similarity coefficient.

The only significant correlation found between the total strength of family-internal groupings and mean accuracy in perceptional experiment for the corresponding target is in the subset of the child-matching task in Finnish: r=0.87, r_s=0.87.

5 Discussion

In the current study, the possibility to perceptually distinguish between the members of the same family and unrelated adult-child pairs in an XAB paradigm appears to be language dependent. Finnish naïve unfamiliar listeners do not attribute adults or children to a particular alien family more than to their own one, but rather they cannot draw any conclusions on perceptual (dis)similarity.

In Russian, the accuracy of answers depends on the target family. Interestingly, the families with the chance-level results are different in the two tasks. For the rest of the targets, respectively, the accuracy is above chance (M=72.5%), and comparable to the results of perceptual identification of twin- and same-gender sibling pairs in voice trios (Decoster et al., 2010; Feiser and Kleber, 2012). The same item can correspond to different mean accuracy for different target families. Thus, rephrasing Rose and Duncan's (1995) conclusion, some voices and some tokens of the same utterance may differ in the identification of the adult-child relationship.

In total, despite some strong family-external similarities, family-internal f0 contour similarities are consistently significantly stronger than family-external in both language groups separately and together. The numerical coefficients of family similarity are not language-specific. However, due to the language-conditioned differences in applied frameworks, it might not be reasonable to compare the strengths of pair similarities between language groups.

In Finnish, both the syllable position in a word and the word position in a sentence have a significant effect on f0 in the proposed five-word three-syllable sentence framework. Unlike the adjacent similarities between syllables, which look rather as a language property, adjacent word similarity represents a gross picture of f0 falling in the sentence and therefore to a certain extent reflects the individual's speech rhythm. Two out of three participating families (H and P) demonstrate strong parent-child resemblance in adjacent word similarities and consistently strong internal similarity in the final coefficient groupings. However, the similarity in the final coefficient groupings of the third family (L) also shows a tendency to be stronger or more consistent than its members' external similarities (especially those of L-parent). Whether the reason of this distinction lies in L-child's older age, smaller parent-child age difference, L-parent's hearing disadvantage (self-reported tinnitus) or other, remains unclear.

In Russian, both position and section have a significant effect on f0 in the proposed three-position three-section framework (changeable number and order of positions in a sentence). The similarities of the sections inside and across positions, as well as of the positions among themselves are believed to reflect the individual's speech rhythm. Parent-child resemblance range from nearly identical in two families (OO and VN), slightly less similar in one family (AL) and showing the greatest dissimilarity in the other (MA). The latter family is also characterized by the weakest internal similarity in the final coefficient groupings, which are, however, in five out of six cases stronger than the family-external similarities of its members.

Interestingly, the parents' family-internal similarities are always stronger or/and more consistent than their similarities to alien children, which is not always the case the other way around. Hence, the individual characteristics of adult's f0 contours pervasively appear in the speech of their children, but children can noticeably demonstrate features that are found in

other adult speakers of the same language, which is most probably reflecting the classic extremes, biology and socialization (Bolinger, 1989) in parent-child intonation similarities.

The accuracy of the Finnish-speaking listeners' performance in the perceptual experiment shows no dependency on the target, nor on the distractor in a trio of voices, albeit family-internal and external similarities vary. The correlation ($r=0.87$, $r_s=0.87$) found between the similarity strength and mean accuracy of answers in the child-matching (always second) task might signal that the participants get used to the material and are attempting to base the decision, which child sounds more like the adult's offspring, on family-internal f0 similarities. However, it seems that either the similarities, as proposed by the current framework, are not prominent enough or the listeners rely on other voice cues.

The results of the perceptual experiment on Russian seem to interestingly reflect the specificity of an XAB discrimination paradigm. Selecting the answer between A and B, listeners in fact make a decision about X. In the parent-matching task, listeners do not choose the parent (A or B) but attribute the child (X) to one of the adults. Albeit AL-family demonstrates high internal similarity in acoustic analysis, AL-parent's external similarities are also strong. Thus, a listener cannot "learn" within the task to map the features exclusively of AL-child to AL-parent and gives more incorrect answers for target AL. The low internal similarity of MA-family does not bring the accuracy for MA-target down because MA-parent's external similarities are weaker. In the child-matching task, on the opposite, a listener cannot "learn" within the task to map the features exclusively of MA-parent to MA-child due to a combination of low internal similarity strength per se and the differences between it and the average external similarity of MA-child. It is also important to note that proposed explanation concerns only the average results of the perceptual experiment. Not all the sentences from the perceptual experiment were acoustically analyzed, which means that they reflect less f0 contour similarities than the selected ones. Half of the non-selected sentences correspond to quite high accuracy results (median 68.9%). Hence, although the f0 contour similarities between Russian-speaking parents and children contribute to identification of family pairs in a trio of voices by non-familiar listeners, the relationship is not linear, may have certain thresholds and involve other voice cues.

6 Conclusion

The current paper presents an attempt to find f0 contour similarities between parents and their young children.

The authors fully acknowledge the limitations of the present study. Analyzed data are limited in their amount, on the one hand, and to read-aloud speech, on the other. The recording scripts do not necessarily reflect naturally occurring utterances in terms of pragmatic, which plays an especially important role in Finnish. The f0 contours are analyzed mostly in their static parameters, nor are other voice cues analyzed as possible contributors to perceptual similarity.

However, the presented findings can be used for the further research on perceptual and acoustic voice similarities between parents and young children or, broader, family members of different age (and gender).

References

Hanna Anttila. 2009. Interrogative intonation in spontaneous Finnish. In V. de Silva and R. Ullakonoja (Eds.) *Phonetics of Russian and Finnish, general description of phonetic systems, experimental studies on spontaneous and read-aloud speech,* pages 167-176. Frankfurt am Main: Peter Lang.

Paul Boersma and David Weenink. 2017. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.33, retrieved 29 September 2017 from http://www.praat.org/

Natalya V. Bogdanova. 2001. *Zhivye foneticheskie protsessy russkoi rechi. Posobie po spetskursu [Live phonetical processes in Russian speech. Course manual]*. Saint Petersburg: Philological Faculty of Saint Petersburg State University.

Dwight Bolinger. 1986. *Intonation and its parts: Melody in spoken English.* London: Edward Arnold.

Dwight Bolinger. 1989. *Intonation and its uses: Melody in Grammar and Discourse.* Stanford, California: Stanford University Press.

David Brazil, Malcolm Coulthard, and Catherine Johns. 1980. *Discourse intonation and language teaching.* London: Longman.

Elena A. Bryzgunova. 1977. *Zvuki i intonatsiya russkoy rechi [Sounds and intonation of Russian speech]* (3rd ed.). Moscow: Russkiy yazyk.

David Crystal. 2003. *A dictionary of Linguistics and phonetics* (5th ed.). Blackwell Publishing.

Viola de Silva, and Riikka Ullakonoja. 2009. Introduction: Russian and Finnish in Contact. In V. de Silva, and R. Ullakonoja (Eds.), *Phonetics of Russian and Finnish. General Description of Phonetic Systems. Experimental Studies on Spontaneous and Read-aloud Speech,* pages 15-20. Frankfurt am Main: Peter Lang.

Frans Debruyne, Wivine Decoster, Annemie Van Gijsel, Julie Vercammen. 2002. Speaking fundamental frequency in monozygotic and dizygotic twins. *Journal of Voice,* 16(4):466-471.

Wivine Decoster and Frans Debruyne. 2000. Longitudinal voice changes: Facts and interpretation. *Journal of Voice,* 14(2):184-193.

Wivine Decoster, Annemie Van Gijsel, Julie Vercammen, and Frans Debruyne. 2001. Voice similarity in identical twins. *Acta Otorhinolaryngologica Belgica,* 55(1):49-55.

Hanna S. Feiser and Felicitas Kleber. 2012. Voice similarity among brothers: evidence from a perception experiment. In *Proceedings of the 21st Annual Conference of the International Association for Forensic Phonetics and Acoustics (IAFPA)*. Santander, Spain.

Michael Fuchs, Jens Oeken, Thomas Hotopp, Roland Täschner, Bettins Hentschel, and Wolf Behrendt. 2000. Similarity of monozygotic twins regarding vocal performance and acoustic markers and possible clinical significance [Abstract]. *HNO,* 48(6):462-469.

Tamas Hacki and S. Heitmüller. 1999. Development of the child's voice. International *Journal of Pediatric Otorhinolaryngology*, 49(1): 141-144.

Pekka Hirvonen. 1970. *Finnish and English communication intonation*. Turku: Publications of the Phonetics Department of the University of Turku 8.

Antti Iivonen. 1977. Lausefonetiikan tutkimuksesta [About the research on phonetics in clauses]. In K. Suomi (Ed.), *Selected papers of the VII Phonetics Symposium Fonetiikan päivät – Turku 1977,* pages 1-14. Turku University, Finland.

Antti Iivonen. 1978. Is there interrogative intonation in Finnish? In E. Gårding, G. Bruce, and R. Bannert (Eds.), *Nordic Prosody: Papers from a Symposium,* pages 43-53. Department of Linguistics, Lund University, Sweden.

Antti Iivonen. 2005. Intonaation käsitteen täsmennystä [Specification of the concept of intonation]. In A. Iivonen (Ed.), *Puheen salaisuudet: Fonetiikan uusia suuntia [Speech secrets: New directions in phonetics],* pages 93-128. Helsinki: Guadeamus.

Patrik N. Juslin and Klaus R. Scherer. 2008. Vocal expression of affect. In J. Harrigan, R. Rosenthal, and K. Scherer (Eds.) *The New Handbook of Methods in Nonverbal Behavior Research,* pages 65-136. Oxford University Press.

Rachel E. Kushner and Corine A. Bickley. 1995. Analysis and perception of voice similarities among family members. *Journal of the Acoustical Society of America 98:* 2936.

Pavel Labutin, Sergey Koval, and Andrey Raev. 2007. Speaker identification based on the statistical analysis of f0. In *Proceedings of IAFPA*. The College of St Mark and St John, Plymouth, UK.

Tatiana M. Nikolaeva. 1970. *Frazovaya intonatsiya slavyansih yazykov [Phrasal intonation of Slavic languages]*. Moscow.

Francis Nolan, Kirsty McDougall, and Toby Hudson. 2011. Some acoustic correlates of perceived (dis)similarity between same-accent voices. In *Proceedings of the 17th ICPhS,* pages 1506-1509. Hong Kong.

Jiska S. Peper, Rachel M. Brouwer, Dorret I. Boomsma, René S. Kahn, and Hilleke E. Hulshoff Pol. 2007. Genetic influences on human brain structure: A review of brain imaging studies in twins. *Human Brain Mapping*, 28:464-473.

Beata D. Przybyla, Horii Yoshiyuki, and Michael H. Crawford. 1992. Vocal fundamental frequency in a twin sample: Looking for a genetic effect. *Journal of Voice*, 6(3):261-266.

Phil Rose and Sally Duncan. 1995. Naive auditory identification and discrimination of similar voices by familiar listeners. *Forensic Linguistics*, 2/1:1-17.

Phil Rose. 1999. Differences and distinguishability in the acoustic characteristics of *hello* in voices of similar-sounding speakers: A forensic phonetic investigation. *Australian Review of Applied Linguistics*, 22(1):1-42.

R Core Team. 2017. R: A language and environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from ttps://www.R-project.org

Eugenia San Segundo and Hermann Künzel. 2015. Automatic speaker recognition of Spanish siblings: (monozygotic and dizygotic) twins and non-twin brothers. *Loquens*, 2(2), e021.

Swapna Sebastian, Anto Suresh Benadict, Geethu K. Sunny, and Achamma Balraj. 2013. An investigation into the voice of identical twins. *Otolaryngology online journal*, 3(2):1-7.

Elaine T. Stathopoulos, Jessica E. Huber, and Joan E. Sussman. 2011. Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age. *Journal of Speech, Language, and Hearing Research*, 54:1011-1021.

Kari Suomi, Juhani Toivanen, and Riikka Ylitalo. 2006. *Fonetiikan ja suomen äänneopin perusteet [The basics of phonetics and Finnish phonology]*. Helsinki: Guadeamus.

Kari Suomi, Juhani Toivanen, and Riikka Ylitalo. 2008. *Finnish sound structure: Phonetics, phonology, phonotactics and prosody.* Oulu: Oulu University Press.

Paul M. Thompson, Tyrone D. Cannon, Narr, K.L., Theo G.M. van Erp, Veli-Pekka Poutanen, Matti Huttunen, Jouko Lönnqvist, Carl-Gustav Standertskjöld-Nordenstam, Jaakko Kaprio, Mohammad Khaledy, Rajneesh Dail, Chris I. Zoumalan, and Arthur W. Toga. 2001. Genetic influences on brain structure. *Nature Neuroscience*, 4(12):1253-1258.

Riikka Ullakonoja, Hanna Kärkkäinen, and Viola de Silva. 2007. Havaintoja venäjän kielen lukupuhunnan sävelkulun ja spontaanin puheen jaksottelun oppimisesta [On learning to perceive pitch in read-aloud speech and spontaneous speech segmenting of Russian]. In O.-P. Salo, T. Nikula, and P. Kalaja (Eds.), *Language in Learning – AFinLA Yearbook*, 37(65), pages 215-231. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AfinLA.

Kristiane M. van Lierde, Bart Vinck, Sofia De Ley, Gregory Clement, and Paul Van Cauwenberge. 2005. Genetics of vocal quality characteristics in monozygotic twins: A multiparameter approach. *Journal of Voice*, 19(4):511-518.

VocaliD, Inc. (2015-2017). https://www.vocalid.co

Natalya B. Volskaya. Aspects of Russian Intonation. In V. de Silva and R. Ullakonoja (Eds.), *Phonetics of Russian and Finnish, general description of phonetic systems, experimental studies on spontaneous and read-aloud speech*, pages 37-46. Frankfurt am Main: Peter Lang.

Melanie Weirich and Leonardo Lancia. 2011. Perceived auditory similarity and its acoustic correlates in twins and unrelated speakers. In *The 17th International Congress of Phonetic Sciences (ICPhS XVII): Congress Proceedings*, pages 2118-2121.

Hadley Wickham. 2009. *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Donghui Zuo and Peggy Mok. 2015. Formant dynamics of bilingual identical twins. *Journal of Phonetics*, *52*, 1-12.

Zvukovaya forma russkoy rechi. Uchebnik po fonetike russkogo yazyka [Sound form of Russian speech. Russian phonetics manual]. 2001-2002. Saint Petersburg: Saint Petersburg State University, Philological Faculty, Department of Phonetics and Experimental Phonetics Lab. Retrieved from www.speech.nw.ru/Manual