<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Logopedics, Phoniatics, Vocology*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

White Sjölander, P. (2001)

Long-term average spectrum (LTAS) analysis of sex- and gender-related differences in children's voices

Logopedics, Phoniatics, Vocology, 26(3): 97-101

<https://doi.org/10.1080/14015430152728007>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-93472>

Long-term average spectrum (LTAS) analysis of sex and gender-related differences in children's voices

Peta White

KTH Voice Research Centre, Speech Music & Hearing, KTH, Stockholm, Sweden

Long-term average spectrum (LTAS) analysis offers representative information on voice timbre providing spectral information averaged over time. It is particularly useful when persistent spectral features are under investigation. The aim of this study was to compare perceived sex of children to the LTAS analysis of their audio signals. A total of 320 children, aged between 3 and 12 years, were recorded singing a song. In an earlier analysis, the recorded voices were evaluated with respect to perceived and actual sex by experienced listeners. From this group, a subgroup of 59 children (30 boys and 29 girls) was selected. The mean LTAS revealed a peak at 5 kHz for children perceived with confidence as boys, and a flat spectrum at 5 kHz for children perceived confidently as girls (whether male or female in actuality).

Key words: children, development, voice, perception, acoustics, spectral analysis, LTAS.

Address for correspondence: Peta White, KTH Voice Research Centre, Speech Music & Hearing, KTH, Stockholm, Sweden¹

BACKGROUND

Whereas adult voice has been the subject of much debate and numerous articles, particularly in recent years, child voice has been regarded traditionally as an area of limited importance. One reason is that children's voice problems have been generally thought to disappear spontaneously during puberty. However, it actually seems more reasonable to assume that voice habits established at an early age may, in common with other human behaviours, persist into adulthood.

According to such sources as Laukkanen (3) and the Yearbook of Statistics Centre of Finland (11), as many as 38% of adults work in professions where voice is regarded as an essential tool of trade. Thus there is, in addition, a socio-economic demand for investigations into the causes as well as extent of

¹ Now Peta Sjölander, KTH, petasj@kth.se

adult voice dysfunction. If there is a chance that some answers might come from childhood, appropriate questions should be asked about this age-group also.

The socio-psychological study of children's voices is also warranted as, for example, it has recently been revealed that a child's personality is an element of strong relevance in chronic hoarseness. And voice quality or style may influence the perception others have of us. For instance, chronically hoarse children were reported by parents and teachers to be talkative and loud voiced. The teachers also added that these pupils were outgoing, open, energetic, less stable and concentrated (5, 6). Where open and energetic might be viewed as positive descriptors, less stable may well not be.

Acoustic study of children's voices, as opposed to adult voices, is further required because the voice organ of a child is not simply a smaller version of the adult voice organ. Indeed, the solutions a child invents to use this organ for speech have long since been a challenge for phoneticians. Unfortunately, due to the high pitches used by children, acoustic analysis is more likely to produce errors in measurement (4, 8) and children are notoriously less reliable than adults as subjects. However, it is necessary to establish the boundaries of normal voice production in the developing child before assumptions can be made with regard to abnormal voice.

Long-term average spectrum (LTAS) analysis has been found to offer representative information on voice timbre. It is characterised by peaks reflecting the average of formant frequencies. This aspect contributes to the perceptual properties of the voice. LTAS analysis provides spectral information averaged over a period of time thereby highlighting the longer term aspects of speech or singing voice production. The method is particularly useful when persistent spectral features are under investigation, features that might not be apparent in shorter samples or single speech sounds. Also, when the sample is long (upwards of 20 seconds) the resulting spectrum is not greatly affected by differences in speech material. This increases confidence in comparisons between speakers and between studies. Earlier studies of 11-year-old children have revealed differences in the centre frequencies of LTAS peaks (9) and formants (10) between sexes indicating the worthiness of further investigation.

AIMS

This study forms part of an ongoing investigation which aims to discover, using acoustic and aerodynamic methods, (a) by what process an immature voice develops into an adult voice; (b) at what age, and for what periods, these changes occur normally; and (c) how development differs between boys and

girls. This article reports on the actual (sex) and perceived (gender) differences between boys' and girls' voices.

METHOD

A total of 320 children, 20 boys and 20 girls in each of eight age-groups (3-4, 5, 6, 7, 8, 9, 10 and 11-12 yrs), were selected from a larger group of over 800 children who had been recorded singing a nursery rhyme (7). A song as opposed to a poem or similar was used in the original study as it was designed to investigate the children's pitching ability. Two songs were specially written or adapted for that study.

In an earlier unpublished analysis, the recorded voices were evaluated with respect to perceived sex and age by 5 listeners who were all experienced teachers of singing to children. With regard to perceived sex, a seven-point scale, ranging from *I'm sure it's a boy* to *I'm sure it's a girl*, with *could be either* as the mid-point option, was used to ascertain each listener's prediction of sex and the level of confidence in his or her decision. The difficulty in classifying the voices correctly varied between age-groups. Analysis of the scores given by each listener provided a subject score of between 7 (= sex identified correctly and with confidence) and 1 (= sex *incorrectly* but confidently identified). A score of 4 indicated that the subject's sex was unclear to the listener.

Fig. 1 presents the mean across-listener scores for the 320 children. As can be seen, many scores are high indicating the listeners were able to predict the sex of a fairly large number of subjects accurately. Most of the low scores are concentrated within the youngest age-groups.

Fig. 1. Mean across-listener scores for 320 boys and girls where 7 is equal to a correct and confident sex identification by all experienced listeners, and 1 is equal to an incorrect but still confident identification.

Fig. 2 shows the trendlines associated with these scores for *boy* and *girl* data. A point of interest here is that the accuracy of sex prediction rose with age for the boys group, but stayed relatively constant with age for girls.

Fig. 2. Trendlines relating to the scores shown in Fig. 1 showing a marked increase in accuracy of identification with increasing age for boys.

For the purposes of the present study, the 20 subjects most often *correctly* and confidently identified as boys, and the 20 most often *correctly* and confidently identified as girls, were selected from this group of 320 ($n = 40$, see Table 1).

Table 1. Mean and median scores and age for 20 boys and 20 girls most often correctly identified by listeners

	n	score	age
<i>mean</i>			
boys	20	6.98	9 yrs
girls	20	6.83	8 yrs
<i>median</i>			
boys	20	7.00	10 yrs
girls	20	6.75	7 yrs

Another criterion of selection was that all children sang the same song. This song lasted approximately 15 seconds when sung as follows:

*There's a fox in a box in my little bed
My little bed, my little bed
There's a fox in a box in my little bed
And there isn't much room for me*

As the children all performed the same song, all sounds, including voiceless sounds, were included in the analysis. The pitch range is one octave and subjects were free to choose the starting pitch.

These subjects are henceforth referred to as ‘high score’ subjects. An LTAS was made of each child’s rendition of the nursery rhyme. A mean was then taken of the data for the 20 boys and another for the 20 girls.

To enable a comparison, subjects with a low score, *ie* of 2.75 or less, were also selected ($n = 19$) and an LTAS was performed in the same way. These children formed a group who were confidently but *in* correctly identified as either boys or girls: *ie* the girls sounded like boys and the boys sounded like girls to the listeners in the earlier experiment. This group is henceforth referred to as ‘low score’ (see Table 2).

Table 2. Mean and median scores and age for 10 boys and 9 girls most often incorrectly identified by listeners

	n	score	age
<i>mean</i>			
boys	10	2.38	5½ yrs
girls	9	1.92	8 yrs
<i>median</i>			
boys	10	2.50	5 yrs
girls	9	2.00	7½ yrs

RESULTS

Fig. 3 shows the mean LTAS for *high score* boys and girls, where a high score resulted from the correct identification of sex in these children by all listeners.

Fig. 3. LTAS of high scoring boys versus high scoring girls (where a high score indicates correct identification of sex by all listeners). A dashed line is drawn at 5 kHz.

There is an immediately observable difference in the curves for these two groups with a peak at around 5 kHz for the boys, indicating that a boy-like sound produced a peak at 5 kHz whereas a girl-like sound produced a relatively flat spectrum at 5 kHz.

To test for the relevance to perception of this 5 kHz peak, the same LTAS analysis was conducted for *low score* subjects. The results, as shown in Fig. 4, demonstrate there is no peak at 5 kHz for the boys, and the girls produced a peak at 5 kHz.

Fig. 4. LTAS of low scoring boys versus low scoring girls (a low score indicates incorrect identification of sex by all listeners). A dashed line is drawn at 5 kHz.

DISCUSSION

Although intriguing, the full relevance and cause of the 5 kHz peak in the LTAS for boys cannot be fully determined at this stage and further analysis is needed before any conclusions can be drawn. However, given that these subjects were consistently and confidently identified (correctly) for sex, the presence or absence of a 5 kHz peak provides at least some evidence of a relationship between perceived sex, as provided by voice quality in the audio signal, and the averaged spectrum.

A corroborating factor is that the peak exists also in the spectra of girls who were *wrongly* but confidently identified as boys.

While the likelihood is growing that the LTAS reflects timbral qualities relating to sex, the causes of the differences seen are, as mentioned above, unknown. For example, the peak might not contain harmonic information and this is an important factor that needs to be determined.

A further consideration is that there may be one or more other factors, present in the LTAS, which influence perception. For example, a relationship may

exist between two or more peaks in the spectrum, or their absence, which is not obvious at this early stage of analysis. Alternatively, there might exist a significant difference in the relative levels of the first and second harmonics which is a factor known to influence perception (1, 2).

Synthesis of a child's voice with and without a 5 kHz peak may be the key to providing valuable information on the effects and relevance to perception of smaller spectrum changes.

One way to further test the perceptual qualities of these voices would be to bandpass filter the samples to, say, 2 or 3 kHz, to see if listeners can detect sex differences with higher frequency information removed, and *vice versa*.

Reliable correlates between these long-term average spectra and perception have yet to be fully established. However, the results presented here go some way towards providing evidence of a link which may be corroborated by further research.

ACKNOWLEDGEMENTS

This research has been funded by the European Commission with a *Marie Curie* Post-doctoral Fellowship www.cordis.lu/improving.

REFERENCES

1. Bickley CA. Acoustic analysis and perception of breathy vowels. In: Working Papers, vol. 2. Cambridge MA: Speech Communications Group, Research Laboratory of Electronics, MIT, 1982. pp. 71–81.
2. Gauffin J, Sundberg J. Spectral correlates of glottal voice source waveform characteristics. *Journal of Speech & Hearing Research* 1989; 32: 556–6.
3. Laukkanen A-M. *On Speaking Voice Exercises*. Doctoral thesis, University of Tampere, Finland 1995.
4. Lindblom B. Accuracy and limitations of sonograph measurements. *Proceedings of the 4th International Congress of Phonetic Sciences*. The Hague: Mouton 1962; 188–202.
5. Sederholm E, McAllister A, Dalkvist J, Sundberg J. Aetiological factors associated with hoarseness in ten year-old children. *Folia Phoniatrica* 1995; 47: 262–78.
6. Sederholm E. *Hoarseness in ten -year -old children: perceptual characteristics, prevalence and Etiology*. Doctoral thesis, Karolinska Institute and KTH, Stockholm 1996.
7. Welch GF, Sergeant DC, White PJ. *Singing Development in Early Childhood: Final Report to the Leverhulme Trust*, London 1994.
8. White PJ. *Acoustic and Aerodynamic Measurements of Children's Voices*. Unpublished doctoral thesis, University of Surrey 1997.
9. White PJ. The effect of vocal intensity variation on children's voices using long-term average spectrum (LTAS) analysis. *Logopedics Phoniatrics Vocology* 1998; 23 (3): 111–20.
10. White PJ. Voice source and formant frequencies in 11-year-old girls and boys. In: White P (editor). *Child Voice*. Stockholm: KTH, 2000. pp. 13–26.
11. Yearbook of Statistics Centre of Finland 1998.

SAMMANFATTNING

Långtidsmedelvärdesspektrumanalys (LTAS) av könsrelaterade skillnader i röstkällan hos barn

Analys med långtidsmedelvärdesspektrum (LTAS) erbjuder representativ information om röstklang. Det visar spektral information medelvärdesbilden över en tidsperiod och är mycket användbart när genomgående drag hos spektrum utforskas. Avsikten med denna studie är att jämföra upplevt kön hos en stor grupp skolbarn med hjälp av LTAS av deras audiosignaler. Totalt 320 barn spelades in när de sjöng en barnsång. Gruppen var fördelad i åtta åldersgrupper mellan tre och tolv år med 20 pojkar och 20 flickor i varje. Vid en tidigare analys utvärderades de inspelade rösterna med avseende på upplevt kön av erfarna lyssnare (Figs. 1 and 2). En hög poäng (max 7) betydde att sångarens kön blev rätt identifierad med stor upplevd säkerhet, och en låg poäng (min 1) betydde att sångarens kön blev fel identifierad men också med stor upplevd säkerhet. Av dessa 320 valdes bara försökspersoner som fick mycket höga poäng ($n = 40$) eller mycket låga poäng ($n = 19$). Dessa två grupper blev uppdelade efter kön. Genomsnittligt LTAS för pojk och flickgrupperna uppvisade en spektraltopp vid 5 kHz för barn som med hög upplevd säkerhet identifierades som pojkar, och ett flackt spektrum för barn som med hög upplevd säkerhet identifierades som flickor, men som kunde vara antingen pojkar eller flickor i verkligheten (Figs. 3 and 4).

YTTHEENVETO

Sukupuolen arviointi lasten äänistä pitkäaikaisspektrumilla

Analyysi pitkäaikaisspektrumilla (long-term average spectrum, LTAS) tarjoaa edustavaa tietoa äänen piirteistä. Se antaa spektraalista informaatiota, joka on muodostettu pidemmän aikajakson aikana, ja on hyvin käyttö kelpoinen kun tarkastellaan pysyviä spektraalisia piirteitä. Tämän tutkimuksen tarkoituksena oli tarkastella miten koettu sukupuoli näkyy LTAS analyysissä, kun tarkastelun kohteena on iso ryhmä koululapsia. Kaikkiaan 320 lapselta nauhoitettiin ääninäyte, kun lapsi lauloi erästä lastenlaulua. Ryhmä oli jaettu kahdeksaan ikäryhmään (3-12 vuotta), kussakin ryhmässä oli 20 tyttöä ja 20 poikaa. Ensimmäisessä analyysissä kokeneet kuulijat arvioivat äänestä koettua sukupuolta (figs. 1 ja 2). Korkea pistemäärä (max 7) kuvasi varmaksi arvioinniksi koettua oikeaa arviointia laulajan sukupuolesta, kun taas matala pistemäärä (min 1) kuvasi varmaksi arvioinniksi koettua väärää arviota lapsen sukupuolesta. Tutkimuksen toiseen vaiheeseen valittiin 320:sta ääninäytteestä koe henkilöiksi vain hyvin korkeat pisteet saaneet ($n = 40$) ja hyvin matalat pisteet saaneet ($n = 19$). Näitä kaksi ryhmää jaettiin sukupuolen mukaan. Keskiarvoistettu LTAS tyttöjä poikaryhmissä näytti

piikkiä 5kHz:n kohdalla niillä lapsilla, jotka oli suurella varmuudella koettu pojiksi ja tasaista spektrumia 5kHz:n kohdalla niillä lapsilla, jotka oli suurella varmuudella koettu tytöiksi. Tulos oli riippumaton siitä, kumpaa sukupuolta lapset todellisuudessa olivat (figs. 3 ja 4).