Acoustic, perceptual and physiological studies of ten-year-old children’s voices

McAllister, A.

journal: TMH-QPSR
volume: 38
number: 1
year: 1997
pages: 075-103

http://www.speech.kth.se/qpsr
Acoustic, perceptual and physiological studies of ten-year-old children’s voices

Anita McAllister

Abstract
The aim of the present investigation was to study hoarseness and its perceptual, acoustic and physiological characteristics in ten-year-old children’s voices.

Fifty-eight children’s voices were perceptually evaluated along 15 voice parameters. The rank ordered means of each parameter revealed a discontinuity in the distribution for all parameters except pitch, breathiness and vocal fry. This discontinuity was used as an operational borderline between normal and deviant voice characteristics. Statistical analysis showed that hyperfunction, breathiness and roughness are the main predictors of hoarseness.

Pitch and intensity ranges of 60 children were recorded in voice range profiles (VRP). Vocal fold status was determined by video-microlaryngoscopy and, when possible, also by stroboscopy. Six children, all boys had vocal nodules. Twenty-five children had incomplete glottal closure. The children had somewhat compressed VRP contours reflecting a more restricted pitch range and dynamics than adults.

The occurrence of register transitions in children’s voices was studied in an experiment. Five voice experts perceptually identified such transitions from a tape made of the VRP recording sessions. On this tape, the stimuli occurred in ascending pitch order. One transition was identified in most voices at a mean fundamental frequency (F0) of A#4. A second transition was identified in four voices at a mean F0 of A#5.

SPL and subglottal pressure were measured at different pitch and loudness levels in nine children. At phonation threshold and at normal conversational loudness the children’s subglottal pressures were similar to those of adult female voices.

Acoustic correlations to perceptual voice characteristics were analysed in two investigations. For the first, six children representing different degrees of hoarseness were selected. The voice sample was running speech. Two perturbation measures were examined but no correlation was found between the perceptual evaluation and these two measures. For the second investigation, 50 children were chosen. The material consisted of sustained vowels. The acoustic measures were period perturbation quotient (PPQ), amplitude perturbation quotient (APQ), and two harmonics-to-noise ratio measures, NNEa and NNEb, operating in different frequencies. The results showed that hoarseness, breathiness and roughness correlated with the acoustic measures PPQ and NNEa at the p≤.001 level of significance. Hyperfunction and instability did not correlate with any of these measurements.

Conclusions: Hoarseness in children’s voices is a stable concept consisting of three main predictors; hyperfunction, breathiness and roughness. Children generally had somewhat compressed VRP recording as compared to adults. Register transitions could be identified in most children’s voices, approximately 25% higher in F0 than for adults. Incomplete glottal closure may be regarded as a normal finding in ten-year-old children. Subglottal pressure values in these children’s voices were similar to those of adult females. In children’s sustained

1 Dissertation presented at the Karolinska Institute, Stockholm, March 14, 1997
Dept of Logopedics and Phoniatrics, Karolinska Institute, Huddinge University Hospital, and Dept of Speech, Music and Hearing, Royal Institute of Technology (KTH), Stockholm
vowel phonation hoarseness, breathiness and roughness correlated with the acoustic measures PPQ and NNEa.

Introduction

The human voice is one of the most important tools in human everyday communication. Indeed, a well-functioning voice is for most people an indispensable apparatus in their everyday life. According to Laukkonen (1995, and personal communication) about 43% of the total labor force in Finland are dependent on a well-functioning voice in their every day work. The languages of the world make extensive and systematic use of the voice potentials in the production of speech sounds, but the voice was probably important to human communication even before language evolved. Also, long before the infant develops language, it expresses important information by means of its voice organ. The voice is obviously of great importance to the expression of human emotions. In human esthetic expression the possibilities of the voice are utilized to the extremes. Coloratura singing, for example, efficiently combines linguistic and emotional expressivity with a virtuosic motor control of respiratory, glottal, and resonatory function.

The complexity challenges characteristics, possibilities and limitations of the voice has been the object of a wide range of human interests, including scientific research. It seems reasonable to assume that much of the essential characteristics of vocal habits are established in childhood. Therefore, the child voice seems to represent a significant aspect which is likely to yield important basic knowledge on the human voice.

This thesis was instigated by some questions regarding a suspected deterioration of children’s vocal health raised by the music teachers at a well known music school in Stockholm, focusing mainly on choir singing (Adolf Fredriks musikskola). The teachers felt that the children applying for the school had gotten hoarser and more dysphonic during the last decade, and that they had a more restricted fundamental frequency (F0) range than previous applicants. In 1991, the Swedish Parliament sent a letter to the National Board of Health and Welfare asking for the opinion of experts regarding this observation. Because of the music teachers’ reports of an increased incidence of hoarseness in children’s voices, this project was initially aimed at analyzing the nature and occurrence of hoarseness as well as possible underlying causal factors. Sederholm (1996a) investigated the prevalence of hoarseness as well as personal and social factors contributing to the vocal health of the ten-year-old children participating in this study. Together we studied the term hoarseness and established its underlying predictors in children’s voices. The present thesis is mainly concerned with the relationship between a set of acoustic and physiological measures and the perceptual characteristics of these children’s voices.

The prevalence of voice deviation in children has been the subject of a number of investigations over the last century. This considerable research effort yielded widely varied results indicating a prevalence ranging between 1% and 80%. This suggests that the methods and the definitions of hoarseness in this research has varied. In fact, in most of these studies hoarseness was never explicitly defined. Among speech pathologists and phoniatricians, hoarseness is considered an ambiguous term composed of several other voice parameters (Sonninen, 1970). Yet, deviant child voices are often described by non specialists as hoarse. A consistent finding, however, is that there seems to be a gender difference, boys typically being hoarse more often than girls.

Usually hoarseness in children’s voices has been assumed to generally disappear during puberty and hence should be of limited relevance (Håkanson & Kitzing, 1984). On the other hand, it has been maintained that child hoarseness may reflect organic disease and that chronic child hoarseness sometimes prompt organic voice disorders (Sédlacková, 1960).

Do hoarse children’s voices and hoarse adult voices share the same perceptual and acoustic properties? Regarding adult voices, the research effort has been substantial. Glottal waveform analysis has been an important method in the attempts to relate the perceptual impression of a voice to measurable acoustic characteristics (Koike, 1973; Horii, 1979; Askenfelt & Hammarberg, 1986; Laver et al., 1986; Laver et al., 1992; Pabon & Plomp, 1992). Different methods have been proposed for measuring the pulse-to-pulse irregularities in amplitude and F0 period time.

The vocal folds of children, however, are not just scaled copies of those of adults. Before mutation, the structure of the tissues connecting the glottic mucosa to the underlying structures is looser than after mutation, as the lamina propria is not fully differentiated into layers in children (Hirano et al., 1983). Furthermore, the ratio between the thickness of the mucosa and
membranous portion of the vocal folds is about 1:2 in newborns and 1:10 and 1:12 in adult women and men, respectively (Hirano et al., 1983). They also reported on developmental changes in the vocalis muscle. Obviously there is also a difference in vocal fold length in children as compared to those of adults.

Perceptual and acoustic properties of children's voices have been the subject of several studies. These perceptual and acoustic investigations have shown that also for children's voices perturbation and other acoustic measures can be productive in discriminating between healthy and pathological voices. It has also been indicated that normal children's voices may typically have higher perturbation values than adults (Steinsapiret al., 1986). Considering spectral characteristics and their possible correlation to voice disorders, harmonics-to-noise-ratio (HNR) is another acoustic measure that has resulted from the pioneering work of Yanagihara (1967). This method compares the harmonic and noise energies in the spectrum. Presently, several different automated methods have emerged measuring this relation in pre-selected frequency ranges (Kasuya et al., 1986; Kikuchi et al., 1986).

The voice range profile VRP (or phonetogram) has been proposed as another easy and clinically useful measurement of voice function. It shows two curves representing the sound levels of softest and loudest phonation as a function of F0 throughout the pitch range. Pabon (1991) showed that the distribution of jitter and shimmer typically varies over the voice range. Thus, the perturbation measures change with both pitch and loudness. According to Pabon & Plomp (1988), information on periodicity is a valuable complement to the regular VRP in discriminating healthy and dysphonic voices. What VRP characteristics can be considered normal and deviant in children's voices?

Another relevant aspect of the voice and especially of the singing voice is register and register transitions. However, its nature has remained unclear in the sense that many terms and definitions exist. Register has been perceptually defined as a "group of like sounds or tone qualities whose origins can be traced to a special kind of mechanical (muscular) action" (Reid, 1983, p 296). Thus the register phenomenon is assumed to be confined to the vocal fold vibration mechanism, thus excluding the relevance of vocal tract resonance. Keidar et al. (1987) studied the perceptual nature of chest and falsetto registers. Their results revealed high interjudge reliability regarding the location of the register transition. Register transitions have also been observed in VRP of choir-singer children (Pedersen et al., 1983; Pedersen et al., 1984). Pedersen et al. (1983) proposed that register transitions could be related to a narrowing in the mid area of the VRP or a local decrease in maximal loudness reflected in the upper contour.

Subglottal pressure (Ps) is closely related to the control of vocal intensity and is also relevant to F0. An increase in Ps tends to be accompanied by an increase of intensity and a rise in F0. How does the difference in vocal fold size and structure reflect on the Ps used by children? For example, do children typically have higher subglottal pressures at the softest possible loudness of phonation, i.e. at the phonation threshold?

Summarizing, the differences mentioned in vocal fold size and morphology between children and adults seem to suggest dissimilarities also with regard to the acoustic and physiologic measurements mentioned above. The overall aim of the present series of studies was to investigate 10-year-old children's voice characteristics in perceptual, acoustic and physiological terms.

**Review of the literature**

**Perceptual voice evaluations**

Perceptual evaluation of voice and voice disorders is a principle tool in clinical practice. The Japanese Society of Logopedics and Phoniatrics developed a perceptual evaluation scale called the GRBAS (Grade, Rough, Breathy, Asthenic and Strained) (Hirano, 1981). This evaluation method won widespread use and is often referred to in the literature. Laver (1980) developed a perceptual evaluation protocol with reference voices. Parameters not normally evaluated in a voice screening are also included in this evaluation protocol, such as larynx position, prosody, pharyngeal constriction and resonatory characteristics, such as nasality. Extensive work on perceptual voice evaluations has been done by Hammarberg (1986), including the study of a large number of perceptual voice parameters used in clinical practice. The perceptual evaluation was also correlated with a set of acoustic measures. A profile frequently used is the FB Wilson Voice Profile that also includes a scale for overall severity similar to Grade in the GRBAS scale (Wilson, 1972). The Buffalo III Voice Profile System is maybe specially aimed at the evaluation of children's voices since it also includes different profiles for group behaviour, voice abuse, speech anxiety and resonance (Wilson, 1987). In all it consists of ten different profiles. The voice screening profile in this system was based on that
suggested by Boone (1973; 1977). The comparison of all these systems for perceptual evaluation of voice characteristics is complicated by the fact that such systems are always limited by the language specific, semantic content of the adjectives used.

**Rating scales**

The most common rating scale in perceptual voice evaluation is the 5- or 7-point equal-appearing interval scale (E AIS) (Wilson, 1972; Hammarberg, 1986; Wilson, 1987; Kreiman et al., 1993). E AIS assumes an equal step-wise increase of the specific voice trait described by the parameter in question. However, there is no evidence that the listener perceives the increase of a voice trait in terms of equal magnitude steps. Rather it seems logical to assume that listeners perceive voice traits along a continuum. Hence, potentially available information is lost in a step-wise rating procedure.

A Visual Analogue Scale (VAS) circumvents this disadvantage. It usually consists of a scale with a 100 mm continuous uninterrupted horizontal line which represents a perceptual parameter. It is often used for measuring a variety of subjective phenomena also including perceived pain (Weavers & Lowe, 1990). Recently this scale has been used in the perceptual evaluation of voices (Kreiman et al., 1993; Södersten & Hammarberg, 1993; Rabinov et al., 1995).

**Voice range profiles (VRP)**

The VRP has been suggested as a potentially valuable aid in the description of both healthy and deviant voices (Coleman et al., 1977; Gramming, 1988; Coleman, 1993; Kotby et al., 1995; Uloza & Šiupšinskie, 1995; Åkerlund, 1996). The method dates back to the thirties when Wolf et al. (1935) described the upper contour of the VRP.

Also children’s voices have been investigated using this simple non-invasive method (Table 1). Children have been reported to have a somewhat elevated lower VRP curve as compared to adult voices and sometimes a more restricted dynamic range (Kotby et al., 1995). The dynamics and the F0 range in choir singer children has been studied by Pedersen et al. (1983), Pedersen et al. (1984) and Pedersen (1993). Böhme & Stuchlik studied 277 children from age 5-14. The children below 7 and above 10 were excluded from the final analysis due to inconsistencies in the VRP registration. Thus, a standard VRP could be calculated for the 112 children between 7 and 10 years of age.

**Table 1. Studies of children’s VRP in healthy and pathological voices.**

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Age</th>
<th>N; females (f), males (m)</th>
<th>Voice types</th>
<th>VRP data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pedersen et al. (1983)</td>
<td>8 - 19</td>
<td>47 (f= 47)</td>
<td>choir singers</td>
<td>Example VRPs</td>
</tr>
<tr>
<td>Pedersen et al. (1984)</td>
<td>”school age”</td>
<td>110 organic and functional pathologies, (distribution of sex not mentioned)</td>
<td>Example VRPs</td>
<td></td>
</tr>
<tr>
<td>Böhme &amp; Stuchlik (1995)</td>
<td>5 - 14</td>
<td>277, 112 between ages 7 -10 (f=67, m=45) normal voices</td>
<td>Standard VRP for ages 7-10, N=112</td>
<td></td>
</tr>
<tr>
<td>Kotby et al. (1995)*</td>
<td>7 - 58</td>
<td>100 (f=50, m=50) normal voices</td>
<td>Mean VRP</td>
<td></td>
</tr>
<tr>
<td>Kotby &amp; Orabi (1995)*</td>
<td>9 - 67</td>
<td>75 (f=32, m=43) organic and functional pathologies</td>
<td>Mean VRP</td>
<td></td>
</tr>
<tr>
<td>Van Tichelt &amp; Heylen (1995)**</td>
<td>6:1-11:8</td>
<td>40 (m=40) normal voices</td>
<td>VRPs of song and speech</td>
<td></td>
</tr>
</tbody>
</table>

* These studies also include adult voices. The number of children in these investigations are not specified.

** In Flemish, reviewed in Heylen et al. (1996).
Larynx, pharynx, and trachea

The morphology of the vocal folds in children differ from that in adults in several respects. Before mutation, the structure of the tissues connecting the glottal mucosa to the underlying structures is looser than after mutation, as the lamina propria is not fully differentiated into layers in children (Hirano et al., 1983). Hirano & coworkers (1983) found that before the age of 4 the vocal ligament did not exist as an identifiable histologic entity. Between the age of 4 - 16 the collagenous and elastic fibers increase in density and become oriented along the length axis of the vocal folds thus forming the deep layer of the lamina propria, or the vocal ligament. The development of the vocal ligament, like the growth of the depth of the vocal folds, appears to continue through puberty. Kahane (1982) found that after the age of 16 adult morphology is present and according to Gray et al. (1993) this three-layered structure could be consistently observed from the age of 15. Each of these distinct layers in the adult vocal fold has specific mechanical properties (Hirano & Kakita, 1985). As mentioned, the ratio between the thickness of the mucosa and membranous portion also varies with age (Hirano et al., 1983).

Hirano & coworkers (1983) also reported on developmental changes in the vocalis muscle. In newborns, the muscle fibers were found to be thin and not fully developed until around the age of 27.

Obviously there is also a difference in vocal fold length in children as compared to those of adults. In newborns, vocal fold lengths between 2.5 and 8 mm have been observed (Table 2). At the age of 10, children’s total vocal fold length, including the cartilaginous portion, is in the order of 13-16 mm for girls and boys respectively (Kent & Vorperian, 1995) (Table 3). This should be compared to the adult total vocal fold length, including membranous and cartilaginous portions, of approximately 21 and 29 mm for adult women and men, respectively (Kahane, 1982). The membranous portion is approximately 12 and 15 mm for adult women and men, respectively (Hirano et al., 1983).

The vocal fold lengths reported by Hirano and colleagues (1983) are clearly shorter than those from the other studies apart from Negus (1962). This discrepancy has been suggested to reflect racial differences or crown-heel length differences (Kent & Vorperian, 1995). Another possibility is different measuring techniques.

The pharynx can be divided into the upper region - the naso- or velopharynx, the mid region - the oropharynx and the lower region - the laryngopharynx. According to Crelin (1973), the total length of the pharynx in newborns is approximately 4 cm. In adults, the total length is approximately 12 cm (King, 1952).

Several studies of tracheal growth have indicated that there are linear relationships between both body weight and height, on the one hand, and tracheal length and cross-sectional area on the other (Butz, 1968; Burrington, 1978). According to Titze (1988), the length of the trachea may influence the fundamental frequency at which register transitions occur.

**Table 2. Different studies of vocal fold length in newborns.**

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Vocal fold length (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negus, 1962</td>
<td>3</td>
</tr>
<tr>
<td>Crelin, 1973</td>
<td>4 - 5</td>
</tr>
<tr>
<td>Kazarian, Sarkissian &amp; Isaakian, 1978 (in Kent &amp; Vorperian, 1995)</td>
<td>4.6</td>
</tr>
<tr>
<td>Tucker &amp; Tucker, 1979</td>
<td>6 - 8</td>
</tr>
<tr>
<td>Hirano et al., 1983</td>
<td>2.5 - 3</td>
</tr>
</tbody>
</table>

**Table 3. Studies of vocal fold length at age 10-11.**

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Vocal fold length (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hirano et al., 1983</td>
<td>9 - 13</td>
</tr>
</tbody>
</table>
Glottal closure
Systematic studies of glottal closure in children are scarce. Björk & Wahlgren (1991) studied glottal closure in 8 boys with mutational voices and found that all had incomplete glottal closure at normal loudness extending past the vocal processus. With increased loudness one boy had complete glottal closure.

Södersten & Lindestad (1990) studied glottal closure patterns in young adults (20-35 years old) and found that 82% of the female phonations and 37% of the male phonations were characterised by a posterior glottal chink. Biever & Bless (1989) studied glottal closure patterns in females with normal voices and found incomplete glottal closure of the posterior part to be a normal characteristic. According to Rammage et al. (1992), a glottal chink extending into the membranous portion may be regarded as pathological.

Acoustic measurements
Normal human voices deviate slightly from perfect periodicity (Rozsypal & Millar, 1979; Klatt & Klatt, 1990). Experience from synthesised speech has revealed that this random variation is a rather important characteristic of the human voice (Hillenbrand, 1987). Thus, there is typically a slight random variation in the voice period time between adjacent voice pulses.

Titze et al. (1988) suggested that, at least in theory, the vocal fold system is potentially chaotic and verified this by constructing a simple recursive glottis model which produced doubling and quadrupling of period time. Similar results have been reported by Lauterborn & Partilz (1988). According to Baer (1979), this variation possibly reflects an inherent neuromuscular condition, viz. the slow rate single-motor-unit twitches in the vocal folds. In most normal voices, the acoustic effect of such irregularities in the voice source is subtle. However, in neurologic disorders these irregularities may be quite pronounced (Ramig et al., 1990). Also the heartbeat has been suggested as one contributor to this apparently random cycle-to-cycle variation (Orlikoff, 1989).

Increased vocal fold mass or stiffness due to nodules or other abnormalities may interfere with the vibratory pattern, particularly if the mucosal wave is affected. As a consequence, deviations from periodicity are often greater in pathological than in healthy voices. Using computer simulation of pathological vocal fold vibration, Ishizaka & Isshiki (1976) found that asymmetric vocal fold vibration yields subharmonic structures in the spectrum and that this produced a rough or creaky sound quality.

In several attempts to establish acoustic correlates to perceptual voice parameters in normal and dysphonic adult voices, the irregularities of the F0 curve have been found to be particularly revealing (Table 4) (see e.g., Koike, 1973; Horii, 1979; Askenfelt & Hammarberg, 1986; Pabon & Plomp, 1988). This jaggedness is frequently measured as the mean dissimilarity in period time between adjacent pitch pulses, and referred to as jitter or F0 perturbation. Other measures have also been used, such as shimmer (cycle-to-cycle variations in amplitude) and harmonic-to-noise ratio (HNR). The latter compares the amplitudes of harmonic and inharmonic spectrum components.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>N</th>
<th>Material</th>
<th>Perceptual parameters</th>
<th>Acoustical parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koike</td>
<td>1973</td>
<td>60 (pathol. and healthy voices)</td>
<td>Vowels</td>
<td>----</td>
<td>Relative average pert. F0 (Jitter)</td>
</tr>
<tr>
<td>Horii</td>
<td>1979</td>
<td>6 (healthy voices)</td>
<td>Vowel</td>
<td>Jitter</td>
<td></td>
</tr>
<tr>
<td>Askenfelt and Hammarberg</td>
<td>1986</td>
<td>41 (pathol. male and fem.)</td>
<td>Text</td>
<td>Averaged voice deviation</td>
<td>7 different perturbation measurements</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(roughness, diplophonia, vocal fry, breathiness, creaky, hyper- and hypofunction, voice breaks, grattings).</td>
<td></td>
</tr>
<tr>
<td>Laver et al.</td>
<td>1986</td>
<td>20 (pathological and healthy)</td>
<td>Vowels</td>
<td>Harsh, hoarse and breathy</td>
<td>Jitter and shimmer</td>
</tr>
<tr>
<td>Pabon and Plomp</td>
<td>1988</td>
<td>26 (13 healthy and 13 pathological)</td>
<td>Vowels</td>
<td>Varying</td>
<td>Jitter, shimmer, high frequency noise, spectral slope</td>
</tr>
<tr>
<td>Karnell et al.</td>
<td>1991</td>
<td>2</td>
<td>Vowels</td>
<td>Not specified</td>
<td>3 different perturbation measurement methods</td>
</tr>
</tbody>
</table>

Table 4. Some studies of voice perturbation measures in healthy and pathological adult voices.
Different F0 perturbation measures have been proposed (Lieberman 1961 and 1963; Hecker & Kreul 1971; Hollien et al., 1973; Koike 1973; Takahashi & Koike 1975; Ludlow et al., 1983; Klingholtz & Martin 1985; Hori, 1975). Pinto & Titze (1990) suggested the use of jitter and shimmer as generic terms, not associated with any particular calculation, but also the use of a complementary strict terminology defining the type of perturbation analysis applied. Titze & Liang (1993) recommended that the F0 extraction be based on the complete waveform. They also emphasized the need for further development of the perturbation measurement and a better understanding of its underlying contributors.

Pabon (1991) showed that the distribution of jitter and shimmer typically varies over the voice range so that the perturbation measures change with both pitch and loudness and Pabon & Plomp (1988) argued that information on aperiodicity is a valuable complement to the regular voice range profile recording in discriminating healthy and dysphonic voices. Still, jitter and shimmer have been found productive in discriminating between healthy and pathological voices (Askenfelt & Hammarberg 1986; Laver et al., 1986; Laver et al., 1992). Askenfelt & Hammarberg (op. cit.) analysed relations between different perturbation measures and perceptual ratings of deviant voice qualities in pathological voices. They found the standard deviation of the relative frequency difference between consecutive periods to yield the highest correlation (r=.66 for females and .76 for males).

Kreiman et al. (1993) pointed out that the perceptual evaluation of voice quality is the standard against which other measurements are compared and evaluated. This seems to be a difficult task, possibly in part due to the fact that the frequency resolution of the human ear varies throughout the frequency range. What can we expect from an acoustic measure of voice characteristics? As mentioned earlier, the perceptual evaluation is dependent on several factors not taken into consideration by the acoustic measurements.

Most acoustic studies of children’s voices originate from the last twenty years. As in the case of adult voices, perturbation and different spectral noise measures have been frequently used (Table 5). Most results suggest that these acoustic measures correlate with dysphonia also for children’s voices. Arnold & Emanuel (1979) compared sustained vowels as produced by children with normal and pathological voices and found deviations in the pathological group with respect to mean F0 and spectral noise level. Oates & Kirkby (1980) found that children with vocal nodules had a higher F0 and a greater F0 variability. Kane & Wellen (1985) noted that perceptual judgments of voice disorder severity in children with vocal nodules correlated significantly with jitter and shimmer measures. Glaze et al. (1988) studied the relations between body height and a set of acoustic voice measures and reported that the voices of taller children typically had higher shimmer and lower HNR values. In a study based on sustained vowels, Steinsapir et al. (1986) reported higher jitter values for children than for adults.

Table 5. Some clinical studies of different acoustic measures in healthy and pathological children’s voices.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Age, Years</th>
<th>N: females (f), males (m)</th>
<th>Material</th>
<th>Perceptual parameters</th>
<th>Acoustical parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arnold and Emanuel, 1979</td>
<td>7.11-10.11</td>
<td>20 m (10 healthy and 10 nodules)</td>
<td>Vowels</td>
<td>Roughness</td>
<td>F0 and spectral noise</td>
</tr>
<tr>
<td>Oates and Kirkby, 1980</td>
<td>10</td>
<td>36 m (12 healthy, 12 breathy 12 nodules)</td>
<td>Vowels</td>
<td>Trained-breathy, breathy-tense, normal</td>
<td>F0, spectral noise, F0 variability, amplitude variability, rel. amplitude</td>
</tr>
<tr>
<td>Hori, 1983</td>
<td>10-12</td>
<td>36 (f=18, m=18 healthy voices)</td>
<td>Text</td>
<td>Not specified</td>
<td>F0 characteristics</td>
</tr>
<tr>
<td>Kane &amp; Wellen, 1985</td>
<td>6.1-11.6</td>
<td>10 (m=6, f=4, all nodules)</td>
<td>Vowel</td>
<td>Harshness/severity (Wilson voice profile system)</td>
<td>Jitter and shimmer</td>
</tr>
<tr>
<td>Steinsapir et al., 1986</td>
<td>3.3-5.11</td>
<td>120 sex not specified (healthy voices)</td>
<td>Vowels</td>
<td>Not specified</td>
<td>Jitter and shimmer</td>
</tr>
<tr>
<td>Glaze et al., 1988</td>
<td>5-11</td>
<td>121 (f=62, m=59, healthy voices)</td>
<td>Vowels</td>
<td>Perceptually normal phonatory quality</td>
<td>Jitter, shimmer and HNR ratio</td>
</tr>
</tbody>
</table>
Vocal register

Register has been perceptually defined as a “group of like sounds or tone qualities whose origins can be traced to a special kind of mechanical (muscular) action” (Reid, 1983, p 296). The register phenomenon is often assumed to be confined to the vocal fold vibration mechanism only, thus excluding the relevance of vocal tract resonances. However, based on evidence from a theoretical model of the voice organ, Titze (1988) suggested that also resonance is relevant to the register phenomenon. According to his theory, register boundaries occur when sub- or supraglottal resonance’s return significant reflections of the glottal pulses to the glottis in an unfavorable phase of their vibration. This implies that the phenomenon of vocal register is dependent on the resonatory systems in the trachea and vocal tract as well as on the vocal fold mechanism. Hence, the F0 at which register transitions occur can be expected to differ depending on tracheal and vocal tract dimensions.

As mentioned, VRP characteristics related to register phenomena in children’s voices have been observed by Pedersen et al. (1983) and Pedersen et al. (1984).

Subglottal pressure

\( P_s \) in cm H\(_2\)O or kPa (kiloPascal) is mostly estimated from oral pressure during p-occlusion recorded by means of a tube inserted in the corner of the mouth. This yields a good approximation of the actual subglottal pressure as derived from tracheal puncture (Hertegård et al., 1995) or a miniature pressure transducer above and below the glottis (Löfqvist et al., 1982).

Titze (1992) has pointed out that the phonation threshold pressure \( P_{th} \), i.e., the \( P_s \) used for softest possible phonation, is relevant to glottal aerodynamics. He also presented a formula for relating \( P_{th} \) to F0. In two articles, Titze (1989, 1992) suggested that \( P_s \) may play a more dominant role in the F0 control when the vocal folds are short and lax. According to Bickley (1991), the vocal folds cease to vibrate at an approximate pressure of 3 cm H\(_2\)O.

Several studies of \( P_s \) in children have indicated that there may be a reversed relation between \( P_s \) and age, that is as age increases, \( P_s \) decreases (Subtelney et al., 1966; Bernthal & Beukelman, 1978; Brown, 1979; Stathopoulos & Weismer, 1985; Stathopoulos & Sapienza, 1993). However, Arkebauer et al. (1967) found that children typically had a lower \( P_s \) than adult speakers. No difference between sexes has been observed for 8-10-year-old children (Stathopoulos & Weismer, 1985; Trullinger & Emanuel, 1983). In a recent investigation, Stathopoulos (1996) found a mean value of 8.7 cm H\(_2\)O for 10-year-old children’s conversational speech. It has been indicated that children with vocal nodules have significantly higher subglottal pressure than their peers (Lotz et al., 1984).

Aims of the present investigations

The present investigations were focused on perceptual, acoustic and physiologic characteristics of 10-year-old children’s voices. The aims were:

- to define the term hoarseness in children’s voices and to find means to identify the hoarse individuals in a sample of normal children (I),
- to identify the children with chronic hoarseness (II),
- to study how hoarseness and other voice characteristics of nonsinger children are reflected in the voice range profile (VRP) (II),
- to investigate children’s vocal fold status and relate it to perceptual and physiological voice characteristics as manifested in the VRP (II),
- to perceptually analyse the occurrence of register transitions and examine their relations to the VRP contours in children’s voices (III),
- to study \( P_s \) and SPL at threshold and at different loudness levels throughout the pitch range in the voices of children aged 8.5-11.5 years (IV),
- to explore the relations between F0 perturbation and perceived hoarseness as well as its three predictors hyperfunction, breathiness and roughness in running speech (V) and
- to analyse the correlation between perceptual and acoustic characteristics of sustained vowel phonation in children (VI).

Methods

Subjects

The project originated from a question regarding a possible deterioration of 10-year-old children’s vocal health raised by music teachers in a music school in Stockholm (Adolf Fredriks musikskola). The subjects were chosen from that age group except in study IV, where there was a span from 8.5-11.5 years of age. A total of 63 children (25 girls and 38 boys) from three schools in the Stockholm area were selected for studies I and II. In studies III and V, sub-
selections from this group of children were made on the basis of different selection criteria.

The nine children in study IV were chosen informally on the basis of their own interest to participate, three girls and six boys. Four of these children, three girls and one boy, had also participated in another unpublished study in which an automated method of recording VRP was used (Pabon et al., manuscript). A schematic overview of the included subjects is shown in Figure 1.

The 50 children in study IV were selected from a group of 205 children so as to represent different degrees of hoarseness (Sederholm, 1996). For a summary of the subjects and some investigation procedures, see Table 6.

Table 6. Subjects in the different studies (f = females, m = males) and methods used.

<table>
<thead>
<tr>
<th>Study</th>
<th>Subjects</th>
<th>Age</th>
<th>Recordings</th>
<th>Speech material</th>
<th>Questionnaires</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>24 f, 34 m</td>
<td>10</td>
<td>Audio-tape, DAT</td>
<td>Connected speech, text and spontaneous</td>
<td>Yes (Growth index)</td>
<td>Voice quality, 15 parameters</td>
</tr>
<tr>
<td>II</td>
<td>24 f, 36 m</td>
<td>10</td>
<td>Audio-tape, DAT x 2 Audimetry Microlaryngo-stroboscopy Voice range profile</td>
<td>Connected speech, text Sustained vowel [i:] Sustained vowel [a:]</td>
<td>Yes (Growth index)</td>
<td>Voice quality, 15 parameters Hearing Laryngeal status Glottal closure Pitch range Dynamic range</td>
</tr>
<tr>
<td>III</td>
<td>6 f, 9 m</td>
<td>10</td>
<td>Audio-tape, DAT</td>
<td>Sustained vowel [a:]</td>
<td>Yes (Growth index)</td>
<td>Register transitions</td>
</tr>
<tr>
<td>IV</td>
<td>3 f, 6 m</td>
<td>8.5-11.5</td>
<td>Audio-tape, DAT</td>
<td>Repetitions of [pa] in softest, normal and loudest phonation</td>
<td>No</td>
<td>Subglottal pressure SPL</td>
</tr>
<tr>
<td>V</td>
<td>1 f, 5 m</td>
<td>10</td>
<td>Audio-tape, DAT</td>
<td>Connected speech, text</td>
<td>Yes (Growth index)</td>
<td>Perturbation</td>
</tr>
<tr>
<td>VI</td>
<td>26 f, 24 m*</td>
<td>10</td>
<td>Audio-tape, DAT</td>
<td>Sustained vowel [a:]</td>
<td>No</td>
<td>Perturbation Normalized noise energy</td>
</tr>
</tbody>
</table>

* Selected from a group of 205 children investigated by Sederholm (1996).
Listeners

Seven speech pathologists, all females and experts on voice disorders, evaluated the voice samples in studies I, II, III, V, and VI. They listened to cassette tapes, on which the voices were individually presented for each listener. Each voice was presented three times. The results of the perceptual evaluation of the 50 voices in study VI are presented elsewhere (Sederholm, 1996). The listeners all had extensive training in perceptual voice evaluations based on the work by Hammarberg (1986). They had a minimum of 5 years experience in the field of voice disorders. However, their experience was limited regarding perceptual evaluation of children’s voices. The perceptual evaluation of the voices in study IV was informal, performed by the two experimenters. To identify register transitions in study III, five experts on singing voice were used. They had extensive experience of singing, three being conductors of children’s choirs, one a singing voice pedagogue, and one a speech pathologist and singer.

Tape recordings and voice samples

**Study I:** Fifty-eight of the 63 children were recorded in acoustically reasonably attenuated rooms in the schools using a Sony TCD-D1 digital audio tape recorder and a Sony ECM-55B microphone. The microphone was mounted on a pair of glasses to ensure a stable and constant microphone distance and to eliminate the risk of air blast. The children were asked first to read a short text and then to retell it in their own words.

**Study II:** Two months later, the children were recorded a second time. This recording included 60 of the 63 children, five of which participated for the first time. The recording was made in a sound treated room (532 x 285 x 270 cm) with an ambient noise level lower than 40 dB above 125 Hz. Each child read a short text and sang a song. All children also made a voice range profile (VRP) recording using the standard 30 cm microphone distance for such recordings. The same recording equipment was used as in study I, complemented by a Brüel & Kjær sound level meter (2215).

**Study III:** The VRP recordings from study II were used. For each recording an edited tape copy was made arranging the tones according to a numbered, ascending pitch order. In this tape, the recordings of five voices were duplicated.

**Study IV:** Three recordings were made of the audio signal and of the oral pressure while the subjects produced repetitions of the syllable [pa:]. These recordings were made in a sound treated booth with the recording equipment outside the booth. The first recording comprised only the softest possible phonation. In the second and third, softest, normal and loudest phonation were registered. Repeated recordings were obtained from four subjects. A SONY ECM 959 DT microphone was used for the audio signal, microphone distance 30 cm. Oral pressure was captured by the pressure transducer included in a Glottal Enterprises MSIF-2 inverse filter equipment. The transducer was connected to a thin plastic tube which the subject held in the mouth corner. The audio and oral pressure signals were recorded on separate tracks on a TEAC RD-200 T PCM data recorder. During the recording, the oral pressure signal was monitored on a SIEMENS 34T minigraph chart recorder to ensure useful registration.

**Study V:** Six recordings from study II were selected which represented different degrees of perceived hoarseness. A sentence consisting predominantly of voiced sounds was used, “men han var för liten, han nådde inte upp” (“but he was too short, he could not reach”, [mən ʰaːn vɑːz ŋeːz ˈlɪtɛn, hæn nəːdə iːn ˈʌp]). The mean duration of this sentence was 3.5 sec.

**Study VI:** Fifty voices selected from a larger material of recordings made in acoustically reasonably attenuated rooms in three schools in different areas of Sweden (Sederholm, 1996). The children read a short text and sustained the vowel [a:] for at least 3 sec. The same recording equipment was used as in studies I and II.

Perceptual analysis

The voice quality of each child was evaluated along a set of perceptual parameters chosen from those suggested by Hammarberg (1986). Hoarseness was added thus making a total of 15 voice parameters (I, II, III, V) (Table 7). Using visual analogue scales (VAS), each parameter was represented by a 100 mm line, the extremes corresponding to no and extremely high occurrence of the trait respectively. In the case of voice pitch, however, a 200 mm long line was used and for register four alternative categories could be chosen, “chest”, “falsetto”, “child voice register” and “other”. The answers were given on a rating sheet, one for each voice.
Table 7. Perceptual voice quality parameters used for the evaluation of children’s voices in different studies.

<table>
<thead>
<tr>
<th>Voice Parameter</th>
<th>Swedish equivalent</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoarseness</td>
<td>Heshet</td>
<td>I, II, III, V, VI</td>
</tr>
<tr>
<td>Breathiness</td>
<td>Läckage</td>
<td>I, II, III, V, VI</td>
</tr>
<tr>
<td>Hyperfunction</td>
<td>Hyperfunktion</td>
<td>I, II, III, V, VI</td>
</tr>
<tr>
<td>Hypofunction</td>
<td>Hypofunktion</td>
<td>I, II, III, V</td>
</tr>
<tr>
<td>Gratings</td>
<td>Skrap</td>
<td>I, II, III, V, VI</td>
</tr>
<tr>
<td>Roughness</td>
<td>Skrovlighet</td>
<td>I, II, III, V, VI</td>
</tr>
<tr>
<td>Voice breaks</td>
<td>Registerbrott</td>
<td>I, II, III, V</td>
</tr>
<tr>
<td>Unstable pitch/quality</td>
<td>Instabil klang/läge</td>
<td>I, II, III, V, VI</td>
</tr>
<tr>
<td>Hard glottal attacks</td>
<td>Härda ansatser</td>
<td>I, II, III, V</td>
</tr>
<tr>
<td>Vocal fry</td>
<td>Knarr</td>
<td>I, II, III, V</td>
</tr>
<tr>
<td>Audible inhalation</td>
<td>Hörbar inandning</td>
<td>I, II, III, V</td>
</tr>
<tr>
<td>Hypernasality</td>
<td>Öppen nasalering</td>
<td>I, II, III, V</td>
</tr>
<tr>
<td>Hyponasality</td>
<td>Sluten nasalering</td>
<td>I, II, III, V</td>
</tr>
<tr>
<td>Pitch*</td>
<td>Röstläge</td>
<td>I, II, III, V</td>
</tr>
<tr>
<td>Register**</td>
<td>Register</td>
<td>I, II, III, V</td>
</tr>
</tbody>
</table>

* The VAS scale for pitch was 200 mm since values above and below neutral could be expected.
** The categories were chest, normal child register, falsetto and other.

In study III, the five listeners were given a test tape with the tones ordered in ascending pitch. The listeners were asked to perceptually identify where register transitions occurred in softest and loudest phonation. Each tone was given a number which was announced on the tape. The listeners marked the transitions on a rating form. Transitions were accepted which were observed by at least two listeners.

Laryngoscopic examination

In study II, sixty children were examined by the late phoniatrician Patricia Gramming, using indirect microlaryngoscopy, and if possible, also stroboscopy. The entire examination could be followed by both the child and the examiner on a Philips color monitor (8802). The examination was recorded on a Nordmende VHS audio-video recorder (V 1005) by means of a Panasonic (WVCD 110-AF) video camera connected to a Zeiss microscope (Opmi 9-FC), and a Brüel & Kjær stroboscope (4914). Lidocain spray (4%) was used for topical anaesthesia of the oropharynx, when necessary. The children produced the sustained vowel [i:] at comfortable pitch and loudness.

The examination included larynx, ear, nose, and throat. The observations were written down on a specific form constructed for this purpose see Appendix B of study II (McAllister et al., 1994). It also included anamnestic data regarding colds and allergies thus allowing an identification of cases of acute hoarseness due to infection.

The video recordings were also examined by another expert phoniatrician who was asked to

describe vocal fold color and shape and to quantitatively evaluate glottal closure according to a form adapted from Södersten & Lindestad (1990), see Appendix C of study II (McAllister et al., 1994).

Audiometry

The children in study II were screened for hearing problems in a sound treated booth using a GSI 17 audiometer. The screening level was set to 20 dB(A).

Voice range profiles (VRP)

VRP recordings were made of 60 children’s voices in study II. The recording procedure followed the recommendations of the Union of the European Phoniatricians (Schutte & Seidner, 1983). However, a flat frequency curve was used for measuring the SPL rather than the dB(A) curve (Gramming & Sundberg, 1988; Pabon & Plomp, 1988). The flat frequency curve is equally sensitive throughout the frequency range and has been shown to facilitate physiological interpretations of the VRP contours (Gramming, 1991).

The subjects were asked to sing at each specific pitch on the vowel [a:] first as softly as possible and then as loudly as possible. A synthesiser, CASIO SA-20, was used for giving the subjects reference pitches. To obtain the desired pitch, it was sometimes necessary for the experimenter to also sing along with the child at the target pitch and then to let the child continue alone. The subjects started at a middle pitch and
continued with higher pitches in ascending order. Then, the lower pitches were recorded in descending order.

**Acoustic analysis**

**Method 1**

For the six voices analysed in study V, fundamental frequency (F0) extraction was made using a variant of the time-domain double-peak-picking method described by Dolansky (1955), and modified by co-author Sten Ternström. The SWELL program was used, developed by Ternström (1992). The input signal was first high- and low-pass filtered. For these children's voices, the filters were set to 150 Hz (+6 dB/octave) and 500 Hz (-18 /octave), respectively. The filtered signal was fed to a peak-tracking circuit that identified the dominant positive and negative peaks and computed the period time. Thus two competing signals representing period-time were extracted, one from the positive peaks of the waveform and one from the negative peaks. Each of the two period-time sequences thus obtained was passed through a 7-point median-value filter. The system chose a period time that derived from either the positive or the negative peaks. For each pitch period, that period time was selected which was closest to its own running median value over the seven most recent periods. This period-time was inverted and output as the F0 value. This procedure yielded a comparatively stable F0 extraction.

Since inflection continuously varies F0 during running speech, it was necessary to prevent the overall pitch trend from influencing the perturbation data. The long-term pitch trend was computed as follows. The raw F0 contour was passed through a digital smoothing filter (20 Hz, -18 dB/octave Butterworth low-pass). Perturbation values larger than 10% were discarded since they tended to reflect pitch tracking errors. Two F0 perturbation measures were used. One was the standard deviation of the relative frequency difference between the raw F0 contour and the pitch trend. The other was the mean of the absolute perturbation, again taking the pitch trend into consideration.

**Method 2**

In study VI, the samples of sustained [a:] produced by 50 voices, were copied on a DAT recording that was acoustically analysed by co-author Seishi Hibi using an automated voice analyser, Rion SH-10. The voice samples consisted of 2.5-3 seconds long portions spliced from the middle portion of the sustained vowel [a:]. Sound level ramps, linear as a function of time and 12.5 ms in duration, were applied to the onset and the decay of the stimuli. Each voice sample occurred three times on the DAT tape.

The Rion SH-10 analyser is a quasi-standard device that has been used in several investigations of healthy and dysphonic voices (Kikuchi et al., 1986; Fex et al., 1994). A detailed technical description of the voice analyser is given by Kasuya et al. (1986). The device performs a running analysis of short (seven F0 periods) intervals of the input signal. Apart from F0 measurement, the analysis included two perturbation measurements based on Koike (1973), period perturbation quotient (PPQ) and amplitude perturbation quotient (APQ), more commonly called jitter and shimmer. PPQ and APQ reflect departures from a running F0 average of period time and amplitude, thus reducing the influence of slow pitch changes. The method also yields a harmonics-to-noise measure called normalised noise energy, or NNE, based on adaptive comb filtering in the frequency domain. It yields an estimate of the mean level of the noise components appearing between the harmonic spectrum partials. In its current version, the device provides two versions of the NNE measure, not included in the description referred to above, the NNEa and NNEb. The only difference is that these measures analyse two different frequency ranges, 0-5 kHz and 1-5 kHz, respectively. For all measurements, all three copies of the same samples were measured and the resulting mean values were used for further correlation analysis.

All these 50 samples of sustained [a:] had been previously evaluated by seven voice experts in a listening test (Sederholm, 1996). In this test, the stimulus order on the tape was individually randomised for each of the seven listeners.

**Subglottal pressure analysis**

The speech sample consisted of repetitions of the syllable [pa:] at different F0 covering a range of 18 to 24 semitones. The pitches were given to the subject either from a Bontempi melotrone instrument or sung by the experimenter, following the same pitch order as in the VRP recordings. Data were collected from nine children in three recordings. In the first, the softest possible phonation was recorded in seven subjects. In the second and third, also normal and loudest phonation were measured in four and two subjects, respectively. Repeated measurements for three of the subjects were obtained for all three loudness level.

Subglottal pressure ($P_g$) was estimated from the oral pressure during p-occlusion. A poly-
ethylene tube, with an inner diameter of 4 mm was inserted in the corner of the mouth. As mentioned earlier, this method has been shown to yield a close correlation to the actual mean subglottal pressure as derived from tracheal puncture (Hertegård et al., 1995) or a miniature pressure transducer below and above the glottis (Löfqvist et al., 1982). The tube was connected to a pressure transducer (Glottal Enterprises). The output and the audio signal were recorded on separate tracks of a TEAC RD 200T FM DAT tape-recorder located outside a sound treated booth measuring 3.5 x 3.3 x 2.2 m. The signals were continuously monitored on a monograph writer to ensure that the P, signals had reasonably flat tops and returned to zero during the vowel sounds. Before and after each session, calibrations of P, and SPL were recorded on the same tape.

Statistical analysis
For the statistical analysis, the SPSS/PC + 4.0 statistics and advanced statistics analysis systems were used (studies I, II). In order to determine the interjudge reliability, Cronbach’s alpha reliability coefficient was computed for the listeners’ perceptual ratings of voice quality (studies I, II and VI) and for the perceptual evaluation of register transitions (study III). To evaluate the intra-rater reliability for the second recording in study II, Pearson’s product moment correlations was computed from the five duplicated voices for each of the seven listeners’ ratings along each of the 14 perceptual voice parameters.

Factor analysis by means of a principal factor and subsequent varimax rotation was used for analysing the factors of relevance to the perception of children’s voices in study I. A multiple stepwise regression analysis was used to identify key parameters underlying rated hoarseness in studies I and II. For the statistical analysis of the relations between the acoustic and perceptual data in study VI, SPSS/PC 5.01 statistics and advanced statistics analysis system were used (Norousis, 1990). Pearson’s pairwise correlations were computed for the interrelations between the four acoustic parameters PPQ, APQ, NNEa, and NNEb, and also for the correlations between the six perceptual parameters and the acoustic measures. A multiple stepwise regression with each of the perceptual parameters as the dependent variable was made to further study how these four acoustic parameters describe the perceptual voice characteristics and also to possibly identify the key acoustic measure of each perceptual parameter.

Results
Study I
Listening experiments with seven experienced speech pathologists revealed that hoarseness in children’s voices was a stable concept. Hoarseness showed high positive correlations with the perceptual parameters breathiness, hyperfunction, roughness, unstable pitch/quality and voice breaks. Statistical analysis revealed that hoarseness is composed mainly of three predictors: hyperfunction, breathiness and roughness; these predictors accounted for 91% of the variance of hoarseness. Interjudge reliability was high, especially for hoarseness. The rank ordered mean rating for each parameter revealed a discontinuity in the distribution for most parameters. A typical example is provided in Figure 2. This breakpoint or elbow, determined by eye, was found slightly below 40 mm. The breakpoint was regarded as the boundary between presence or absence of voice deviation. According to this criterion 24% could be regarded as being hoarse.

Study II
The second recording of the children in study I yielded a breakpoint in the rank ordered means at 40 mm, approximately (Figure 3). This corresponded to a hoarseness occurrence of 23%. The children with means above these breakpoints on both occasions were regarded as chronically hoarse. The prevalence of chronic hoarseness in this group of children thus came to 14%. All these eight children were boys.

Fifty out of 60 children could complete a VRP recording. Clear differences were found in the VRP registration between chronically hoarse and non hoarse children as well as between the voices of children with and without vocal nodules and glottal chinks. A typical example pertaining to the children with vocal nodules is shown in Figure 4. For comparison, data from a control group consisting of 24 voices without perceptual and physiological deviations is shown in the same graph. As compared with female adults, the children in the control group had a somewhat restricted dynamic range in terms of an elevated lower and a lower upper contour (Figure 4). However, for the boys with
Figure 2. Rated mean hoarseness values for the 58 voices in study I plotted in rank order. Note the marked discontinuity at the hoarseness value of 35 mm.

Figure 3. Rated mean hoarseness values for the 60 voices in study II plotted in rank order. Note the marked discontinuity at the hoarseness value of 40 mm.
mutational voices the SPL values for loudest phonation were similar to those of female adults.

Laryngoscopy

Laryngoscopic examination was feasible in 51 of the 60 children and stroboscopy could be performed on 32 children (Table 8). Six children had bilateral vocal nodules, three of whom were regarded as chronically hoarse. Of the remaining three voices two participated only in the second recording and hence, chronic hoarseness could not be established in these cases. One child with bilateral vocal nodules was classified as non-hoarse. Four of the children with vocal nodules had incomplete closure in the hourglass shape.

In 25 cases the original visual examination revealed incomplete glottal closure in terms of chinks of shapes 2 - 4 (Figure 5). Three of these children had incomplete closure all along the vocal folds. In four cases of incomplete glottal closure, stroboscopy could not be performed. Thus, the incidence of incomplete closure came to 49% (25 out of 51) from the laryngoscopic examination and 65% (21 out of 32) of the children participating also in the stroboscopic evaluation. The 25 cases were evaluated also by another phoniatrist using video recordings of the examinations. From the video recordings the diagnosis incomplete glottal closure was confirmed in 14 of these 25 cases. In the perceptual evaluation, nine of these 14 children were judged as free of hoarseness, hyperfunction,

Figure 4. Averaged VRP for 5 children with vocal nodules (filled diamonds, lines), a control group of 24 children, 13 girls and 11 boys, with normal glottal status and no perceived hoarseness, hyperfunction, roughness or breathiness (open circles) and 10 nonsinger women with normal voices (open triangles, data from Gramming, 1991). FO is expressed as percentages of the normalized total range.

Table 8. Results from laryngoscopic examination with and without stroboscopy.

<table>
<thead>
<tr>
<th></th>
<th>Complete glottal closure</th>
<th>Incomplete glottal closure all along the vocal folds</th>
<th>Incomplete glottal closure along the posterior half or two thirds</th>
<th>Hourglass chinks</th>
<th>Inspected but closure not evaluated</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro-laryngoscopy</td>
<td>11</td>
<td>3</td>
<td>22</td>
<td>4</td>
<td>11</td>
<td>51</td>
</tr>
<tr>
<td>Stroboscopy</td>
<td>7</td>
<td>1</td>
<td>20</td>
<td>4</td>
<td>-</td>
<td>32</td>
</tr>
</tbody>
</table>
breathiness and roughness. The view of the vocal folds was superior during the examination as compared to that in the video, according to the examiner Patricia Gramming.

Eleven children had complete glottal closure according to the initial examination, seven of these cases were also confirmed by stroboscopy. Eleven children could not sustain phonation long enough to perform stroboscopy.

Audiometry revealed one case of conductive hearing loss of one ear. Normal hearing above 20 dB was established in the remaining 59 children.

**Study III**

Five voice experts listened for register transitions in the VRP recordings of 15 voices and five of which were duplicated on the test tape. The interjudge reliability as determined by Cronbach’s alpha was satisfactory (Table 9). The low alpha value for the second transition in the lower contour was due to few ratings. The mean intrajudge reliability, calculated from the duplicated voice samples, was .72 according to Pearson’s correlation coefficient. One register transition could be detected in both soft and loud phonation in almost all of the 15 analysed voices. The mean F0 of this transition was 472 Hz (=A#4). A second transition was observed at a mean F0 of 902 Hz (=A#5) in four voices, two controls and two mutational. The mean F0 value of the first transition is higher than that typically found in adult female voices. The F0 value of the first transition seems to be in accordance with Titze’s (1988) model for vocal registers, based on influence of sub- and supraglottal resonances. No relation was found between register transitions and discontinuities in the VRP contours.

Table 9. Interjudge reliability coefficients for ratings of the location of the first and second register transitions in the lower and upper contour.

<table>
<thead>
<tr>
<th>Perceived transition, number</th>
<th>Cronbach’s alpha, a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper contour; transition 1</td>
<td>.77</td>
</tr>
<tr>
<td>Lower contour; transition 1</td>
<td>.88</td>
</tr>
<tr>
<td>Upper contour; transition 2</td>
<td>.80</td>
</tr>
<tr>
<td>Lower contour; transition 2</td>
<td>(.47)</td>
</tr>
</tbody>
</table>

**Study IV**

The mean subglottal threshold pressure for the nine children examined was similar to those predicted for adult female voices. However, in the lowest F0 range subglottal pressure values were higher than those predicted by Titze’s equation (1994). The scatter of data points was considerable, for a given pitch varying between 4 -5 cm H2O.

At normal conversational loudness and at loudest level of phonation the children’s subglottal threshold values were between 2-4 and
4-8 times the subglottal threshold values predicted by Titze’s equation, respectively. Many children’s subglottal pressures tended to increase with the square root of fundamental frequency. In soft phonation, the children’s overall SPL values were somewhat lower than those previously measured in phonotragram recordings of both children and adult females. This discrepancy may be due to the difference in experimental procedure used in the phonotragram recordings and in the present phonation threshold pressure recordings. For a doubling of subglottal pressure, the average SPL increase was found to be 10.5 dB.

**Study V**

Two perturbation measures were used on six children’s running speech. The influence of pitch inflection on these measures was minimised by comparing individual period time with the overall trend in the F0 contour. The children were selected so as to represent different degrees of hoarseness. No clear correlations were found between perceived hoarseness in these children’s running speech and the two perturbation measures used, the standard deviation of F0 perturbation and the mean of the absolute perturbation. Low correlation’s were found between these two perturbation measures and two of the predictors of hoarseness, hyperfunction and breathiness. The third predictor roughness showed a correlation of 0.46 and 0.44 with the SD of the perturbation and the mean of the absolute perturbation, respectively.

Perturbation values are generally used for sustained vowels while in this study children’s running speech was investigated. Applying the same analysis method on four adult voices showed that the perturbation values for running speech were clearly higher than those for sustained vowels, even though the influence of pitch inflection in running speech had been eliminated.

**Study VI**

The sustained vowel phonation of 50 children, selected so as to represent different degrees of hoarseness, was acoustically analysed by co-author Hibi using a commercially available automated voice analyser, RION SH-10. Our results revealed a correlation between the perceptual parameters hoarseness, breathiness and roughness and the acoustic measures PPQ and NNEa at the p≤.001 level of significance, although the correlation coefficients varied between r=.734 and r=.456. Breathiness correlated with all acoustic measures. No correlation was found between hyperfunction and instability and any of these acoustic measures. APQ correlated only with breathiness, r=.403. Thus, the acoustic measures considered seem related to breathiness, hoarseness and roughness but not to hyperfunction and instability. By and large, these results are similar to those previously found for adult voices.

**Discussion**

**Perceptual voice evaluation**

Perceptual analysis of voice quality is always closely related to language. The semantic content of the terms used vary in different languages. Also, features that are accepted as idiomatic in one language are regarded as deviant in another. For example, in about 10% of the world’s languages a breathy voice is used as a phonemic marker and breathy phonation is also known to be used as a phonemic contrast in many different language groups (Ladefoged, 1983; Ladefoged & Maddieson, 1996). This clearly complicates international comparisons of e.g. the prevalence of hoarseness and poses the question whether or not international standards regarding normality and deviation of voice can be established.

Also other factors affect the perceptual evaluation of a voice. Sonninen & Sonninen (1976) found that e.g., age, social and cultural factors are relevant. Thus, the perception of normality depends also on our previous experience. On the other hand, data from our questionnaire to the teachers and parents indicated that experts and parents made similar judgements regarding hoarseness in children’s voices (Sederholm et al., 1995). Thus, our speech pathologists seemed to possess a judgement of voice deviation in the perceptual evaluation of children’s hoarseness that was common to that of the teachers and the parents. This indicates that hoarseness, a term generally avoided by specialists, may be quite useful in the communication between clinicians and patients. However, a similar agreement would not be expected for other perceptual voice parameters; also, trained listeners usually have higher interjudge reliability than naive listeners (Bassich & Ludlow, 1986; Deal & Belcher, 1990).

Traditionally, perceptual judgements and the associated reliability of the emerging results are central in the study of voice quality and this thesis makes extensive use of such evaluations. In studies I, II, III, V and VI the results were based on or compared to perceptual evaluations made by trained voice clinicians, all female. In study IV, the perceptual evaluation was more
impressionistic based on the judgement of two experienced listeners, the experimenters. Study III involves the perceptual rating of register transitions made by five singing voice experts. The grouping of the subjects in this study was based on the perceptual and laryngoscopic evaluations made in study II complemented by data on growth rate collected by Sederholm and coworkers (Sederholm et al., 1995).

Recently Kreiman et al. (1996) argued that perceptual analysis is unreliable, especially for essentially normal voices for which they found an interjudge variability of approximately 20%. Higher degrees of agreement were observed for clearly normal and clearly deviant voices. This is in accordance with our results, showing that the interjudge variability was greatest around the break point in the rank order graphs in studies I and II; it does not seem surprising that interjudge variability is greatest for voices possessing degrees of a particular voice trait that are near the limit of deviance, as different listeners are likely to apply somewhat different definitions of this limit.

Given this background the interjudge reliability in studies I and II was remarkably high, especially considering that no anchor stimuli were used and no training session proceeded the evaluations. The highest interjudge reliability was observed for the parameter hoarseness in both studies (Cronbach alpha value = .92). This finding contrasts the findings of Kreiman et al. (1993), possibly because of the common educational background of the members of our listening panel. Furthermore, our results are in agreement with previous Swedish studies where adult voices were perceptually evaluated by a similar group of experts. This agreement is interesting in view of the fact that reference voices and training sessions were used in these studies, as opposed to our investigation (Hammarberg, 1986; Södersten, 1994).

It is interesting that the interjudge reliability was highest for the hoarseness parameter and somewhat lower for the more specific parameters hyperfunction, breathiness, and roughness. Thus, it seemed easier to judge hoarseness. This suggests the possibility that hoarseness may be a worthwhile analysis parameter to use in studies of subtle voice characteristics that are difficult to perceive in terms of more specific parameters.

A multiple regression analysis of the responses in study I, showed that hoarseness had three main predictors, viz., breathiness, hyperfunction and roughness. The rank ordered means for three of these four perceptual parameters revealed a clear discontinuity in the distribution which was used as the borderline between normal and deviant voice quality. No clear break was observed in the rank ordered graph for breathiness. This break point appears to reflect a categorical perception of the voice trait, discriminating between normal and deviant. Such a categorisation is typical in the perception of complex phenomena, for example in the categorical perception paradigm of phoneme perception (Abramson & Lisker, 1970). It has been argued that, with regard to phonemes, these perceptual discontinuities are attributed to the presence of natural boundaries distinguishing complex acoustic phenomena (Miller et al., 1976). It would thus be interesting to further study the voices immediately below and above the breakpoint in rank ordered mean graphs of each voice trait, forcing our listeners to make a judgement of normality or deviation. Relating these responses to the previous qualitative perceptual ratings would add information about perceptual evaluations and also shed new light on the exact degrees of a voice trait accepted as normal by individual listeners.

Break points in rank plots was observed not only in the results of the perceptual evaluation but also in the distribution of two of the acoustic measures analysed in study VI, viz., PPQ and APQ. The break points in the distribution of acoustic data did not correspond to those in the perceptual evaluation. Still, the existence of discontinuities in acoustic measures supports the assumption that the break points in the perceptual evaluation did not exclusively reflect a phenomenon specific to perception.

Intra-judge reliability was studied using five duplicated voices in study II. However, this low number of duplicated voices caused statistical difficulties due to many zero ratings for some parameters that were apparently absent in the chosen voices. Although this resulted in a low variation, the mean Pearson correlation coefficient was still rather high, .81. The listeners had particular difficulties with the parameters breathiness and vocal fry. This may reflect that these voice traits are regarded as normal characteristics in children’s voices. In adult female voices, breathiness has been shown to be a normal voice characteristic (Henton & Bladen, 1985; Södersten & Lindestad, 1990) and vocal fry may be a normal voice characteristic in English speaking males (Henton & Bladen, 1988). It is possible that vocal fry and breathiness are regarded as normal characteristics also in Swedish children’s voices.
Voice range profiles

Voice range profiles have, especially in Europe, been a popular diagnostic tool, but the reproducibility has sometimes been questioned. This has been studied in investigations concerning the day to day variation of VRP data and possible training effects (Coleman et al., 1977; Frank & Donner, 1986; Gramming et al., 1991; Coleman, 1993). The influence of the experimenter on the test results of the subjects is another possible source of variation (Titze et al., 1995). Coleman et al. (1977), Awan (1991), and Gramming et al. (1991) have reported test re-test mean intra-subject variations of less than 2-5 dB. Thus, softest and loudest possible phonation seem to yield reasonably stable values. This supports the assumption that physiological properties of the larynx are reflected in the VRP.

More specifically, it has been assumed that the VRP contours are related to the structural properties of the vocal folds. The lower contour is likely to reflect the willingness of the mucosa to vibrate at low driving pressures whilst the upper contour would depend on the capacity of the vocalis muscle to cope with high pressures (Gramming & Sundberg, 1988; Kitzing, personal communication). As compared to adults, the upper contour of our control group was lower; this may reflect a restricted ability of the vocalis muscle to resist the high pressures required for loud phonation, especially at high pitches. The lower contour of all children was elevated as compared to that of adults; this may suggest that their vocal fold structure demands higher subglottal pressure to vibrate. Another explanation is that children’s muscular and sensory control is not as developed as in adults causing them to use higher subglottal pressures than required to achieve sustained phonation. Also, higher phonation threshold pressures have been predicted for shorter than for longer vocal folds (Titze, 1992). Furthermore, different criteria regarding acceptable phonation may contribute to this difference. The upper VRP contour of the mutational voices revealed an approximation to adult voices upper contour.

The advantage of a computerised recording procedure, providing the subject with visual feedback has been pointed out by many authors (Pabon et al., manuscript). The benefit of automatic recordings systems also entails the advantage of studying subjects who are unable to sustain a given pitch. In an unpublished study, 12 children were tested using such an equipment, developed by Pabon (1991). The results from that investigation indicates that the F0 and dynamic ranges of children are larger than previously reported (Pabon et al., manuscript). For example, the lower contour was about 10 to 15 dB below those observed with a manual recording procedure. These results are similar to those observed for the phonation threshold in study IV.

VRP recording offer data on pitch range. Our results in study II for nonsinger children was 24 semitones. This value is in rather good accordance with earlier findings. Weinberg (1914) found a mean normal pitch range of 15-16 semitones for a group of both singer and nonsinger children, Flatau & Gutzmann (1908) observed a mean pitch range of 21 semitones in 50 ten-year old children, Hartlieb (1957) reported a pitch range of 24-30 semitones in 937 subjects out of a total of 1301, and Blatt (1983) reported a mean pitch range of 26 semitones for five choir singing boys 9-10 years old. In comparing voice pitch ranges it seems important to distinguish between the true physiological range and a range consisting only of tones that the subject and/or the experimenter regards as “acceptable”. The former is presumably larger, at least in nonsinger subjects. It is possible that with additional practice some of the children in our group could have been able to extend the upper pitch limit of their FO range by a number of semitones. For instance, several of the children had never before sung in their falsetto register and felt uncomfortable with the sound of their voice in that frequency range.

Glottal closure

Glottal chinks were observed during laryngeal examination in 25 of the 51 children by co-author Gramming, study II. Twenty-one cases were also confirmed by stroboscopy. From a video recording of these sessions a second phoniatrician was able to confirm incomplete closure in 14 of these 25 cases. Interestingly, none of these 14 voices were not judged as breathy in the perceptual evaluation. No more than seven of the 51 children had complete glottal closure as verified by stroboscopy.

The incidence of glottal chinks in these 51 children may seem remarkably high. However, Björk & Wahlgren (1991) observed incomplete glottal closure in all the eight male mutational voices examined. Posterior glottal chinks was found normal in adult females (Biever & Bless, 1985) and Södersten & Lindestad (1990) reported such chinks in 86% of the phonation of female adults.
Register transitions

Three major categories of vocal register are consistently mentioned in the literature, vocal fry (pulse) register, modal (chest) register, and falsetto (Titze, 1988; Hirano 1981). Titze proposed a classification into two types of transitions, the periodicity transition between vocal fry and chest register and the timbre transition between chest register and falsetto.

Our expert listeners were able to consistently detect register transitions in 14 of the 15 children’s voices. The inter- as well as the intrajudge reliability was satisfactory. This indicates that child voices typically exhibit at least one clear register transition.

The first register transition was thus identified in all voices but one, mean F0 472 Hz (=Bb4). This is about 3 semitones higher than that found by Wurgler (1990) and Pedersen et al. (1984). However, in some voices the register transitions occurred at a lower F0 in soft phonation than in loud. Pedersen et al. (1983) also observed that some boys with mutational voices were unable to produce certain pitches at the register transition, thus producing an interruption of the VRP contours. Neither of the two mutational voices in our material showed any such contour interruptions.

A second transition was found in four voices only, in both subjects with mutational voices and in two controls. The restricted pitch range of the children with deviant voices did not include the higher F0 values at which this second register transition occurred. The mean F0 of this second transition was 807 Hz (=Ab5) for the controls and 1003 Hz (=B5) for the mutational voices. The latter is somewhat higher than what was observed by Wurgler (1990). On the other hand, it is in reasonable accordance with findings of Flatau (1905) who found a transition around 1048 Hz (=C6) in some pubertal girls voices when they changed into a ”fourth register with a wood instrument timbre”.

Acoustic measures

Study V failed to demonstrate any correlations between hoarseness and its predictors, on the one hand, and the two perturbation measures investigated, on the other. By contrast, study VI revealed a significant correlation (p<0.001) between hoarseness, breathiness and roughness and PPQ and NNEa; breathiness correlated with all acoustic parameters while no correlations were found for hyperfunction and instability.

The lack of correlation between perceptual evaluation of hoarseness and acoustic perturbation measures in study V may be due to several factors. First, it should be remembered that the number of voices was no more than six, thus making any general conclusions tentative. This would be the main reason for the lack of correlation between perceptual and acoustic measures. Second, the material analysed in study V was running speech while in study VI it was sustained vowels. Running speech as opposed to sustained vowels represents a great number of contiguous vowels and should hence offer a representative material also in an acoustic analysis (Scherer et al., 1995).

Klingholz (1990) found running speech to yield more reliable HNR data than sustained vowel phonation in the discrimination between normal and pathological voices. Still, most authors have arrived at the opposite conclusion that sustained vowels are more appropriate for acoustic analysis of perturbation, as such material better meets the demands raised by method of analysis (Koike, 1973; Horii, 1979; Murry & Doherty, 1980; Laver et al., 1986; Pabon & Plomp, 1988). An important reason would be that F0 variations, stress and phonetic context in spoken language tend to distort the perturbation measures example; in study V we used the F0 trendline to minimise the influence of such distortions. However, also supralaryngeal obstructions of the airway in running speech tend to disturb the transglottal airflow, thus producing F0 perturbations (Bickley & Stevens, 1986; Stevens, 1991). Short-term disturbances of the F0 contour are not likely to be taken into consideration by a trendline.

With regard to children’s voice characteristics most attempts to identify acoustic correlates suggest that perturbation measures, derived from sustained vowels, represent relevant information (Arnold & Emanuel, 1979; Oates & Kirkby, 1980; Kane & Wellen, 1985; Glaze et al., 1988). The correlations found in study VI between perceptual and acoustic measures support this assumption.

The correlations found in study VI were rather moderate. However, perfect correlations between perceptual and acoustic voice measures cannot be expected for several reasons. For instance, previous experience, differing perceptual relevance of various parts of the signal, both associated with the limitations of our hearing system affect how a human listener perceives a voice, as was discussed in section Perceptual Evaluation, above.

A problem with the aperiodicity measures is the sensitivity to “inappropriate” factors such as microphone type and placement (Titze & Winholdt, 1989) and, particularly for high-pitched voices, the sampling frequency (Titze et
al., 1987). Karnell et al. (1991) compared different perturbation measures and demonstrated that also other factors such as window duration and method of F0 extraction could greatly affect the resulting jitter values. It seems that higher correlations between perceptual and acoustic measures could benefit from improved methods of analysing F0 perturbation.

In study VI, we found that hoarseness, breathiness and roughness correlated with a HNR measure, the NNEa. Corresponding correlations have been found also for adult voices (Yanagihara, 1967; Kasuya et al., 1986; Yumoto et al., 1984). HNR analysis is mostly run by comb filtering in the frequency domain. According to de Krom (1993) a comb filtering in the cepstrum domain should be preferable, as this method will detect also the perceptually relevant sub-harmonics.

**Voice samples**

Above we discussed the relative appropriateness of sustained vowels as opposed to running speech as the material for an acoustical analysis of voice characteristics. However, in the discussion regarding the usefulness of these two types of speech material representativity is also important to consider.

As was just mentioned, sustained vowels, rather than running speech, are mostly recommended for the purpose of the acoustic analysis of perturbation. On the other hand, Pabon (1991) showed that the distribution of jitter and shimmer typically varies over the voice range, so that perturbation values changes with both pitch and loudness, thus casting doubts as to the representativity of a sustained vowel. Furthermore, Södersten & Hammarberg (1989) found that most perceptual parameters were more prominent in running speech as compared to sustained vowels in adult voices. Similar observation were made by Askenfelt & Hammarberg (1986), although in cases of severe voice pathology they found sustained vowel phonation more representative. According to Fitch (1990), reading a text rendered more stable voice samples as compared to sustained vowel phonation and running speech. Thus, in our material of essentially normal voices, the perceptual characteristics could be expected to be less prominent in sustained vowel phonation.

The perceptual judgement of children’s voices were similar for sustained vowels and running speech; Sederholm (1996) found that, with regard to hoarseness, breathiness, roughness, and hyperfunction, expert listeners made comparable perceptual judgements of 50 children’s running speech and sustained vowels. However, most voice characteristics, apart from roughness, were perceived as more pronounced in running speech as compared to sustained vowel phonation. This possibly indicates that roughness in a sustained vowel is regarded as more deviant than when present in running speech.

De Krom (1994) argued that much of the perceptually important information is present in the onset and decay of the vowel. However, such transients were excluded in study VI in order to equalise the conditions for perceptual evaluation and the acoustic measures. It seems possible that the listener to a large extent forms an opinion about the condition of a voice on the impression of voice onset and decay. However, the acoustic measurement does not make such considerations, but rather gives equal weight to the entire speech segment analysed. In an attempt to equalise the analysing conditions these possibly important segments were removed.

Summarising, in perceptual analysis of children’s voices running speech seems preferable for reasons of representativity. On the other hand, roughness and breathiness has been found to be more readily observable in sustained vowels.

**Subglottal pressure**

The mean phonation threshold pressures observed in our group of children in study IV were somewhat lower when compared to typical values previously reported for children’s soft phonation (Stathopoulos & Sapienza, 1993; Subtelney et al., 1966; Bernthal & Beukelman, 1978; Brown, 1979; Stathopoulos & Weismer, 1985; Stathopoulos ET, personal communication). This could be due to the fact that the children in our investigation were asked to phonate as softly as possible rather than merely soft. However, the threshold pressure values obtained for our group of children were rather similar to those observed for nonsinger adult female voices (Akerlund & Gramming 1994; Gramming 1989) although our data had a greater scatter.

Our $P_{th}$ data deviated slightly from Titze’s (1992) predictions for the lowest F0 region. Similar discrepancies were found for adult female voices (Akerlund & Gramming 1994). Furthermore, discrepancies were found regarding the overall F0 dependence. Titze’s equation is based on a number of physiological parameters, such as mucosal wave velocity in the vocal fold cover, vocal fold thickness and length and mean damping of the tissue vibration (Titze, 1989,
1992, 1996). Also, the level of hydration has been found to affect $P_{th}$ (Verdolini-Marston et al., 1990). It is possible that children and adults differ in all these respects.

In male patients with non-organic dysphonia, significantly higher $P_{th}$ values have been found as compared to normal voices (Gramming, 1989), and high $P_s$ values have been observed in children with vocal nodules (Lotz et al., 1984). In our investigation, one child with acute hoarseness exhibited a clearly elevated $P_s$ values, and his mean $P_{th}$ values were almost twice those of the other subjects and also increased more steeply with $F_0$. This probably reflected an increased stiffness in his vocal fold cover.

The mean $P_s$ value of 5.9 cm H$_2$O for normal conversational loudness is similar to that reported by Statthopoulos & Weisman (1985) for a group of children from 4-8 years of age. It is also similar to values reported for adults at normal conversational loudness and $F_0$ (Fant 1960; Subtelney et al., 1966; Bernthal & Beukelman, 1978; Holmberg et al., 1988; Hertegård & Gauffin, 1991; Hertegård et al., 1995). However, Statthopoulos (personal communication) recently found a higher mean value of 8.7 cm H$_2$O for 10-year-old children’s conversational speech, possibly reflecting cultural and other differences between the subject groups.

A doubling of $P_s$ yielded a mean increase of 10.5 dB in our group of children while an increase of 16 dB was previously reported for 8-year-old children (Statthopoulos & Sapienza, 1993). It is possible that differences in the experimental conditions account for this discrepancy. On the other hand, for adults Ishihaki (1964) and Fant (1982) found a doubling of $P_s$ to result in about 9 dB increase of intensity. Thus, our results suggest that the $P_s$ values are similar for children and adults with respect to phonation threshold, conversational speech and the SPL obtained for a doubling of $P_s$.

As $P_s$ is the main physiological control parameter for vocal loudness, it is also relevant to consider the associated SPL values. In softest phonation, most of our SPL values were low as compared to previous phonetogram data for children (Böhme & Stuchlik, 1995), and also for adult females (Gramming, 1991). They were also low as compared to those obtained for the lower phonetogram contour in study II, again possibly reflecting differences in task. In phonetogram studies, the subjects sustain an isolated vowel at constant pitches and sound levels, while in study IV they repeated a CV syllable; it does not seem unlikely that a higher $P_s$ is needed for sustaining a vowel. The recording procedure is also likely to affect the SPL values. For example, Pabon and coworkers (manuscript in preparation) measured the VRP using an automatic recording procedure that provided visual feedback. For a group of 12 children, they observed SPL values for the lower VRP contour that were similar to our data observed for the phonation threshold in study IV.

Subglottal pressure has also been found relevant to register function. In an experiment using excised larynxes, Berry et al. (1995) found shifts in $P_s$ to be associated with register transitions. However, we cannot determine if such a relationship exists also in children’s voices, as no $P_s$ data were recorded in our investigation of registers (study III). Also, to investigate links between $P_s$ shifts and register phenomena, study IV would need to be complemented by a perceptual evaluation.

Hyperfunction, obviously associated with excessive $P_s$, has been described as one of the main characteristics of hoarseness in children’s voices (Sedlácková, 1960; Nemec, 1961; Wilson, 1987). Studies I and II revealed that hyperfunction is one of the main predictors of hoarseness. Furthermore, Amorosa et al. (1990) suggested that voice problems in children may indicate deficits in fine motor control and coordination and Crary (1993) suggested that also sensory deficiencies in the oral region are often associated with hyperfunctional voice use. Panabacker (1975) found that five children out of eight with chronic hoarseness had articulatory difficulties. In therapy of children with voice problems, it has been observed that children with voice disorders also have an impaired sensory feedback. Appropriate voice use raises high demands on the regulation of $P_s$ which is likely to depend on the sensory proprioception also in the oral region. Thus there are several indications suggesting that it would be worthwhile to examine the oral motor function in children with voice disorders.

**Gender differences**

Several of our findings seem to vary between gender. With regard to incidence of chronic hoarseness all cases were boys (study II). However, in previous investigations boy-to-girl ratios of 3:2 or 3:1 have been reported (Flatau & Gutzman, 1908; Weinberg, 1915; Baynes, 1966; Silverman & Zimmer, 1974, Yairi et al., 1974; Toohill, 1975). The extreme prevalence of boys found in our study were probably not due to language but rather to random factors. In a study of 205 Swedish children, Sederholm (1996) found a ratio of 3:2 between boys and girls with respect to chronic hoarseness, thus corroborating previous findings.
With respect to breathiness the occurrence was somewhat greater for girls, while the boys had clearly more hyperfunctional voices and also a higher incidence of roughness (study VI). A corresponding gender difference was observed in the acoustic analysis (study VI).

Regarding subglottal pressures only three girls were included in the study (IV) thus preventing general conclusions. Still, it is interesting to note that the girls had somewhat lower $P_{th}$ values than the boys.

Outlook

The main aim of the present work was to expand our knowledge of children’s voices, thus improving the basis for management of children with voice disorders. An improved description of children’s voice characteristics and voice function is likely to promote also the understanding of adult voices. The high incidence of hyperfunctional dysphonia in both boys and male adults may be interpreted as an indication that detrimental phonatory habits are established during childhood. If so, early intervention could possibly prevent such voice disorders in the adult population.

There are also other reasons for paying attention to child hoarseness. A hoarse, dysfunctional voice is likely to influence the child’s self perception and it also affects listeners’ evaluation of personality traits in a pejorative direction (Muma et al., 1968; Ruscello et al., 1988; Sederholm, 1996). In addition, a hoarse voice may also prevent a child from pursuing a musical/singing interest. It seems clearly worthwhile to investigate various longterm effects of hoarseness in a longitudinal study. Thereby, the perceptual, physiological as well as acoustic documentation of voice characteristics gathered in these studies should offer a starting point.

In a clinical evaluation, laryngoscopy and perceptual evaluation is routine. Our results suggest that also other factors should be taken into consideration. In particular, $P_{th}$ seems a worthwhile parameter. By means of the method used in study IV normative data could be assembled for children’s $P_{th}$ values, and these values should be compared to those of hoarse children.

Conclusions

The present investigations of ten-year-old children’s voices rendered the following conclusions:

- Visual analogue scales are productive in differentiating between essentially normal and deviant voice quality.
- Hoarseness is a stable concept. It consists of three main predictors: hyperfunction, breathiness and roughness.
- VRP analysis appears to be a useful method for the evaluation of children's voices showing clear differences between chronically hoarse and non-hoarse children.
- Children in general seem to have somewhat restricted dynamic range as compared with adults.
- Incomplete glottal closure may be regarded as a normal finding in children’s voices.
- Children’s subglottal threshold pressures are similar to those predicted for adult female voices.
- A doubling of subglottal pressure yields an average SPL increase of 10.5 dB. This value is similar to that observed in adults.
- In softest phonation, the children’s overall SPL values were somewhat lower than those previously measured in VRP recordings of both children and adult females.
- One register transition can typically be detected at 472 Hz (=A#4) mean F0 in both soft and loud phonation. In some voices, a second transition occurs about one octave higher.
- In sustained vowel phonation the perceptual parameters hoarseness, breathiness and roughness correlated with the acoustic measures PPQ and NNEa. Breathiness correlated also with APQ and NNEb.

Acknowledgements

I wish to express my sincere gratitude to all friends and colleagues at the Department of Speech Pathology, Danderyd Hospital, who in many ways supported and facilitated my work and also tolerated my sometimes very disparate behaviour. I also want to thank the staff at the Department of Speech, Music and Hearing, KTH, for their generosity and helpful attitude.

In particular I wish to thank

Johan Sundberg, Department of Speech, Music and Hearing, KTH, my supervisor and co-author, for the generous sharing of his vast scientific knowledge, outstanding supervision, patience, never-failing support and friendship.

Britta Hammarberg, Department of Logopedics and Phoniatrics, Huddinge University Hospital, my supervisor, for her professional guidance, constructive suggestions and great generosity.

Elisabeth Sederholm, Department of Logopedics, ENT, Umeå University Hospital, my co-
author and colleague, for her skill, reliability and
great sense of humor.

Patricia Gramming, my co-author, for her
enthusiasm, inspiration and for her great
empathy during the laryngoscopic examinations
of the children, and for her friendship. I miss her.

Sten Ternström, Department of Speech,
Music and Hearing, KTH, my co-author, for
invaluable help with acoustic analysis, generosity
and enjoyable discussions.

Seishi Hibi, Department of Otolaryngology -
Head and Neck Surgery, Kurume University,
Japan, my co-author, for great help with acoustic
analysis.

Jan Dalkvist, Department of Psychology,
University of Stockholm, for expert guidance in
statistical analysis, great discussions and
personal support.

Peter Pabon, Institute for Sonology, Royal
Conservatory, Den Haag, the Netherlands, for
interesting research collaboration, many fruitful
discussions and friendship.

Björn Fritzell and Aatto Somninen, professors
and phoniatricians emeriti, for generously
sharing their vast knowledge and clinical experience
and for their great sense of humor.

Kerstin Wahlström-Ohlsson, Royal University
College of Music, Stockholm, for good
advice and knowledgeable comments regarding
register transitions.

Eva Borell, Britt Hedin, Eva Ilhre, Elisabeth
Lundström, Åslög Rosén, Maria Södersten and
Helena Wisniewska, speech pathologists and
voice experts, for their skill and endurance in the
perceptual evaluations of these children’s voices.

Bror Samuelsson, Birgitta Samuelsson, Åslög
Rosén, Gunnel Fagius and Kerstin Wahlström-
Ohlsson, singing voice experts, for their skill and
sharp ears in the identification of register
transitions.

Jenny Iwarsson, Department of Speech,
Music and Hearing, KTH and Department of
Logopedics and Phoniatrics, Huddinge University
Hospital, for help with the preparation of
tapes for perceptual and acoustic evaluation, for
helping with coding of data and for friendship
and many laughs.

Nils Malmsten, Department of Psychology,
University of Stockholm, for great help with
statistical analysis and for opening the door to
the magic world of computers.

My family, in particular Bob, my husband,
for his never-failing support and patient revisions
of my English throughout the work, and to our
children Linda and John for just being and
tolerating all the ups and downs.

Friends, for personal support.

Last, but certainly not least, the children, who
so courageously participated, not only in
recordings of reading and singing but also in the
"uncomfortable" laryngoscopic examination just
because we asked them to.

This work was supported by grants from the
Axel and Margaret Ax:son Johnson’s Foundation
and the Karolinska Institute.

The Patricia Gramming Memorial Foundation
for the contribution to the printing of this thesis.

References

Abramson AS, Lisker L (1970). Discriminability
along the voicing continuum: Cross-language
tests. Proceedings of the Sixth International
Congress of Phonetics, Prague. Academia.

problems in children with unintelligible speech as
indicators of deficits in fine motor coordination.
Folia Phoniatrica, 42: 64-70.

Arkebauer HJ, Hixon TJ, Hardy JC (1967). Peak
intraoral air pressure during speech. J Speech
Hea Res, 10: 196-208.

levels and roughness severity ratings for vowels
produced by male children. J Speech Hear Res,
22: 613-626.

Askenfelt AG, Hammarberg B (1986). Speech
waveform perturbation analysis: A perceptual-
acoustical comparison of seven measures. J Speech
Hear Res, 29: 50-64.

characteristics of untrained versus trained vocal

Baan T (1980) Vocal jitter; a neuromuscular
Transcripts of the eight symposium on the care of
the professional voice, New York: The voice
foundation, 19-24.

Bassich CJ, Ludlow CL (1986). The use of
perceptual methods by new clinicians for assessing

Baynes RA (1966). An incidence study of chronic
hoarseness among children. J Speech Hear Dis, 2:
172-176.

Benthal JE, Beukelman DR (1978). Intraoral air
pressure during the production of /p/ and /b/
by children, youths and adults. J Speech Hear Res,

Berry D, Herzel H, Titze I, Story B (1995). Bifurca-
tions in excised larynx experiments. National
Center for Voice and Speech, Univ of Iowa,
Denver Center for the Performing Arts, Univ of

computer model of a larynx. In: Gauflin J &
Hammarberg B (eds). Vocal Fold Physiology.
Acoustic, Perceptual and Physiological Aspects of
Voice Production, San Diego: Singular Publishing
Group, 37-44.

Bickley CB, Steens KN (1986). Effects of a vocal-
tract constriction on the glottal source: experi-

98
m ental and modelling studies. *J Phonetics, 14*: 373-382.


Flatou TS (1905). Sur un quatrième registre musical de la voix des jeunes filles dans l’age de puberté. *Archives International de Laryngologie*.


Hammarberg B: Personal communication.

Hammarberg B, Gauffin J (1995). Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological


Kitzing P. Personal communication.


a joint meeting of Acoustic Society of America and Acoustic Society of Japan, dec. Hawaii.


Ladefoged P, Maddieson I (1996). The sounds of the world languages. Cambridge, Massachusetts, USA: Blackwell Publisher Inc.


Laukkanen A-M. Pesonal communication.


Pabon JPH, McAllister A, Sederholm E, Gramming P, Sundberg J. Dynamics and voice quality aspects of children’s voices as manifested in computer phonotograms. Manuscript for publication.


Sederholm E (1996). Hoarseness in ten-year-old children’s voices, perceptual characteristics, prevalence and etiology. Doct Diss., Dept of Logopedics and Phoniatrics, Karolinska Institute, Huddinge University Hospital and Dept of Speech, Music and Hearing, Royal Institute of Technology (KTH), Stockholm.

Sederholm E. Perceptual evaluation of connected speech and sustained vowels in children’s voices. (Under publication).


