Dept. for Speech, Music and Hearing
Quarterly Progress and
Status Report

Acoustic measurements and
perceptual evaluation of
hoarseness in children's
voices

McAllister, A. and Sundberg, J. and Hibi, S. R.

journal: TMH-QPSR
volume: 37
number: 4
year: 1996
pages: 015-026

**KTH Computer Science
and Communication**

<http://www.speech.kth.se/qpsr>

Acoustic measurements and perceptual evaluation of hoarseness in children's voices

Anita McAllister¹, Johan Sundberg² and Seishi R Hibi³

¹. Dept of Speech Pathology, Danderyd Hospital, S-182 88 Danderyd, Sweden

². Dept of Speech, Music and Hearing, KTH, Box 70014, S-100 44 Stockholm, Sweden

³. Dept of Otolaryngology - Head and Neck Surgery, Kurume University School of Medicine, 67 Asahi-machi, Kurume 830, Japan

Abstract

According to a previous investigation of children's voices hoarseness was found to be a stable concept consisting of three main predictors: hyperfunction, breathiness, and roughness. The present investigation analyses the relation between these perceptual voice characteristics, complemented by instability and gratings (high pitched roughness), and four acoustic measures, pitch perturbation quotient, amplitude perturbation quotient and two types of normalised noise energy measures as proposed by Kasuya and coworkers. The analysis was carried out on a material of vowels sustained by 50 ten-year-old children with non pathological voices representing different degrees of hoarseness. Significant correlations were found between hoarseness, breathiness, and roughness, on the one hand, and the frequency perturbation quotient as well as one of the harmonics-to-noise measures on the other.

Introduction

In a previous pilot study we investigated the relation between perceived hoarseness and two measures of fundamental frequency perturbation in the running speech of six children (McAllister et al., 1996). The speech material was identical for all voices. The measures examined were the standard deviation of the relative perturbation and the mean of the absolute perturbation. In the computation of these measures, the influence of intonation was eliminated. The perceptual evaluation showed clear differences between the voices with regard to *hoarseness* and its three predictors *hyperfunction*, *breathiness* and *roughness*. However, no clear correlations were observed between these perceptual parameters and the perturbation measures.

This lack of correlation may be due to several different factors. First, the number of voices in our investigation was no more than six. Second, the material analysed was running speech which can be assumed to be more representative for the voices than sustained vowels but also to affect the outcome of acoustic measurements. Askénfelt & Hammarberg (1986) found that sustained vowel phonation was representative of voice deviation only in cases of severe voice pathology, whereas in other cases, running speech rendered more representative voice samples. Regarding the perceptual judgment of children's voices

Sederholm (1996) found that, with regard to *hoarseness*, *breathiness*, *roughness*, and *hyperfunction*, expert listeners made similar perceptual judgements of 50 children's running speech and sustained vowels. However, the magnitude and order of these parameters were not identical in the two ratings. Similar findings were made by de Krom (1994).

Sustained vowels are often recommended for the purpose of the acoustic analysis of perturbation. Murry & Doherty (1980) analysing five healthy and five severely pathological voices (malignancy of the larynx) found that sustained vowels were more appropriate than running speech for a correct classification based on acoustic measures. Perturbation measures from sustained vowels and from running speech differ for many reasons. It has been suggested that a supraglottal constriction of a voiced continuant impedes the airflow, reduces transglottal pressure drop, and perturbs the vocal fold vibrations (Bickley & Stevens, 1986; Stevens, 1991). Thus, such constrictions can be expected to affect perturbation measures. Moreover, Pabon (1991) showed that the distribution of jitter and shimmer typically varies over the voice range, so that perturbation values change systematically with both pitch and loudness.

Most attempts to identify acoustic correlates of children's voice characteristics suggest that perturbation measures derived from sustained

vowels represent relevant information. A number of acoustic investigations of children's sustained vowel production have shown positive results. Arnold & Emanuel (1979), Kane & Wellen (1985) and Oates & Kirkby (1980), analysed sustained vowels as produced by children with normal and pathological voices and found divergent acoustic data for the pathological group as compared to the normal voices.

The correlation between perceived voice deviation and spectrum characteristics has been investigated by several researchers. In 1967, Yanagihara reported on a close correlation between perceived hoarseness and the subjective estimate of noise visible in the second formant region as well as above 3 kHz. Kasuya et al. (1986) developed an automatic method for estimating such noise components in sustained vowel phonation. They analysed a group of 238 voices, 64 with normal glottal status and 174 with vocal pathologies classified on the basis of laryngeal findings. Their method successfully categorised 187 cases, or 79%. Yumoto et al. (1984) showed that hoarseness got a significantly higher correlation with two spectrographic measures, harmonic-to-noise and spectrogram analyses, than with jitter.

Summarising, with regard to children's voices sustained vowels seem to yield reasonably representative voice samples with respect to hoarseness and certain other voice traits (Sederholm, 1996). Also, perturbation measures seem more reliable when derived from sustained vowels than from running speech. The present study aims at analysing the correlation between perceptual and acoustic characteristics of sustained vowel phonation in 10-year-old children. The study concerns six perceptual voice parameters and four acoustic measures chosen on the basis of previous investigations.

Method

Subjects

In a project aimed at determining the prevalence of hoarseness in 10-year-old children's voices, a material of 205 recordings was collected (Sederholm, 1995). The voices represented three different Swedish dialects, none of which uses

phonatory manner as a regional marker (Hammarberg, personal communication). Seven voice experts perceptually analysed all voices with regard to hoarseness and a set of five other voice characteristics (*hyperfunction, breathiness, roughness, instability* and *gratings*). The mean hoarseness value was then computed for each voice. For the present study, 50 of these voices were selected, 26 of boys and 24 of girls. The selection criteria was that their voices were evenly distributed along a scale of rated mean hoarseness values, see Figure 1 (Sederholm, 1996). None of these 50 voices were identical to any of the 6 voices analysed in our previous investigation of the relations between perceptual and acoustic voice characteristics (McAllister et al., 1996).

Tape recordings

The voices of the children were recorded in acoustically reasonably attenuated rooms in the schools using a Sony TCD-D1 DAT tape recorder and a Sony ECM-55B microphone. The microphone was mounted on a pair of eyeglass-frames to ensure a stable and constant microphone distance, and to eliminate the risk of air blast. Each child read a short text and was also asked to sustain the vowel [a:] for at least 3 seconds. In the present study, only the sustained vowels will be considered.

Perceptual evaluation

For the perceptual evaluation 2.5 - 3 seconds long portions were 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 test tape. The stimulus order on the tape was individually randomised for each of the seven listeners. In these sustained vowels the voices had a mean fundamental frequency level of 270 Hz, varying between 215 - 364 Hz. This is only slightly higher than the 262 Hz reported by Eguchi and Hirsh (1969) as the mean speaking fundamental frequency for 10-year-olds.

Figure 1. Mean rated hoarseness for the 50 selected voices plotted in rank order.

The perceptually evaluated voice material was acoustically analysed by co-author SH using a commercially available automated voice analyser, a slightly modified version of Rion SH-10. 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, two harmonics-to-noise measures called normalised noise energy, NNEa and NNEb, analysing the frequency range of 0-5 kHz and 1-5 kHz, respectively. Also, fundamental frequency was measured in Hz. PPQ and APQ reflect departures from a running F0 average of period time and amplitude, thus reducing the influence of slow pitch changes. The NNE measures are based on adaptive comb filtering in the frequency domain and yield estimates of the mean level of noise components appearing between the harmonic spectrum partials.

The method has been described in detail elsewhere (Kasuya et al., 1986; Kikuchi et al., 1986). It was developed for analysing sustained vowel phonation and has been found to successfully identify not only severe vocal pathologies but also milder pathologies such as vocal nodules in adult voices (Kasuya et al., 1986) and to confirm treatment results (Fex et al., 1994).

For all measurements, three repeated computations were made using the resulting mean values for further analysis and correlations. To evaluate the reproducibility of the acoustic

measures, three copies of each voice sample were analysed, and the standard deviation of each measure of the same voice sample was computed. Table 1 shows, for each of the five acoustic measures, these standard deviations averaged across all 50 voices. The standard deviations reveal a good reproducibility.

Table 1. Standard deviations for the three repeated computations of the same voice samples averaged across the 50 voices for all five acoustic measurements.

	PPQ %	APQ %	NNEa (dB)	NNEb (dB)	F0 (Hz)
SD	0,018	0,052	0.18	0.22	0.35

Results

For the statistical analysis of the relations between the acoustic and perceptual data SPSS/PC 5.01 statistics and advanced statistics analysis system were used (Norusis, 1990). Pearson's pairwise correlation was computed.

In a multiple stepwise regression analysis of the perceptual evaluation of these 50 voices' sustained vowel phonation, three predictors of *hoarseness* were identified that together explained 89% of the variance of *hoarseness*. *Roughness* accounted for 66%, followed by *breathiness* and *hyperfunction* accounting for 14 and 9%, respectively. These results are reasonably similar with regard to the selected voice parameters for sustained vowels and connected speech. In the connected speech of these voices,

the statistical analysis selected the same three predictors but with *hyperfunction* as the most prominent variable explaining 73% of the variance of hoarseness followed by *breathiness* and *roughness* accounting for 9 and 1%, respectively (Sederholm, 1996). The listeners perceived the sustained vowels to be more unstable but less hoarse, a result that does not seem surprising considering the different vocal tasks.

Table 2 shows the interrelations between the acoustic parameters in terms of Pearson's two-tailed correlation coefficients and the corresponding levels of significance. All four measures showed significant degrees of correlation varying between $r=.335$ and $.736$. Note that four correlations reached significance at the highest $p\leq.001$ level. Correlations between perturbation and harmonics to noise measures have been observed previously (Wolfe et al., 1995; Yumoto et al., 1984).

Table 3 shows the correlations between the six perceptual parameters and the four acoustic measures. Figures 2, 3 and 4 show the distributions and the regression lines for the statistically most significant correlations between acoustic and perceptual parameters. *Breathiness* correlated with all acoustic measures. For three of the tested acoustic measures, the correlation was at the $p\leq.001$ level of significance, the highest being with NNEa followed by PPQ and NNEb and also APQ, although at the lower level of significance.

Hoarseness and *roughness* were both found to correlate with PPQ and NNEa at the $p\leq.001$ level of significance. *Gratings* correlated with PPQ, but at the lower level of significance. *Hyperfunction* and *instability*, on the other hand, did not correlate with any of the acoustic measures. Among the acoustic measures, APQ and NNEb correlated only with *breathiness*. PPQ and NNEa correlated with three perceptual parameters, *hoarseness*, *breathiness* and *roughness*, at the highest level of significance. In addition, PPQ also correlated with *gratings*, at $p\leq.05$.

The correlation between perceptual and acoustic parameters was also analysed with sex as the dependent variable. In Table 4, the results are shown for girls only. Here *breathiness* was the only parameter reaching the $p\leq.001$ level of significance with PPQ and NNEa. *Breathiness* also correlated with APQ at the lower level of significance. The correlation between *hoarseness* and PPQ and NNEa was lower, $p\leq.05$. No other parameters correlated at the selected levels of significance for the girls.

Table 5 shows the corresponding correlations for the boys. Here both PPQ and NNEa correlated with *hoarseness*, *breathiness* and *roughness*. NNEb correlated with *hoarseness* and *breathiness*. The level of significance varied. Only *breathiness* reached the $p\leq.001$ level of significance with three of the four acoustic measures used; PPQ, NNEa and NNEb.

Table 2. Pearson's two tailed correlation coefficients for the four acoustic parameters together with their corresponding levels of significance, ** $p\leq.001$, * $p\leq.05$.

	APQ		PPQ		NNEa		NNEb	
	r	p	r	p	r	p	r	p
APQ	1	-						
PPQ	.478	.000**	1	-				
NNEa	.574	.000**	.736	.000**	1	-		
NNEb	.335	.017*	.418	.003*	.598	.000**	1	-

Table 3. Correlation coefficients (r) and associated probability levels (p) between the four acoustic measures and the six perceptual parameters for all 50 voices. Significant correlations are indicated by stars, ** $p\leq.001$, * $p\leq.05$.

	PPQ		APQ		NNEa		NNEb	
	r	p	r	p	r	p	r	p
Hoarseness	.598	.000**	.155	.282	.551	.000**	.221	.123
Hyperfunction	.087	.549	-.202	.159	-.124	.393	-.256	.068
Breathiness	.618	.000**	.403	.004*	.734	.000**	.443	.001**
Roughness	.544	.000**	.004	.977	.456	.001**	.146	.313
Instability	.123	.395	.134	.356	.174	.227	.003	.986
Gratings	.393	.005*	-.006	.968	.243	.089	.013	.928

Table 4. Correlation coefficients (r) and associated probability levels (p) between the four acoustic measures and the six perceptual parameters for the 26 girls. Significant correlation's are indicated by stars, ** $p \leq 0.001$, * $p \leq 0.05$.

	PPQ		APQ		NNEa		NNEb	
	r	p	r	p	r	p	r	p
Hoarseness	.490	.01*	.300	.137	.500	.009*	.115	.576
Hyperfunction	-.056	.784	-.136	.507	-.244	.230	-.333	.097
Breathiness	.714	.000**	.459	.018*	.776	.000**	.357	.074
Roughness	.294	.145	.121	.558	.384	.053	.039	.848
Instability	.222	.275	.262	.196	.206	.313	-.034	.868
Gratings	.242	.235	.023	.910	.229	.260	-.051	.805

Table 5. Correlation coefficients (r) and associated significance levels (p) between the four acoustic measures and the six perceptual parameters for the 24 boys. Significant correlation's are indicated by stars, ** $p \leq 0.001$, * $p \leq 0.05$.

	PPQ		APQ		NNEa		NNEb	
	r	p	r	p	r	p	r	p
Hoarseness	.591	.002*	-.003	.989	.570	.004*	.495	.014*
Hyperfunction	.073	.736	-.392	.058	-.106	.624	-.182	.394
Breathiness	.677	.000**	.425	.039*	.766	.000**	.768	.000**
Roughness	.545	.006*	-.191	.372	.478	.018*	.369	.076
Instability	.040	.854	.002	.995	.113	.598	.095	.659
Gratings	.394	.057	-.149	.486	.168	.433	.121	.575

In order 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 a multiple stepwise regression was made with each of the perceptual parameters as the dependent variable. The results are listed in Table 6. For the perceptual variables *instability* and *hyperfunction*, the analysis failed to select any acoustic variables. Nor did these two variables correlate with any of the acoustic variables, see Table 3. Only *breathiness* reached an explained variance above 50% with the acoustic parameter NNEa of 54%. For all other parameters, the explained variance was clearly below 50%. Most equations only proceeded one step thus indicating that further analysis would not add to the explained variance.

In a previous investigation of children's voices, each perceptual parameter was rank ordered according to ascending mean from a perceptual evaluation (Sederholm et al., 1993). For most parameters, a marked discontinuity could be observed in the distribution. This discontinuity was used as a borderline between normal and deviant voices. Parallel to that investigation, the mean values of the acoustic parameters in the present study were rank ordered in figure 2. A clear discontinuity can be observed in the rank order of PPQ for the last three voices. A somewhat smoother change of direction can be observed for APQ. No clear discontinuity can be observed in the rank order means of NNEa and NNEb. The voices at the top of the rank order of the acoustic parameters are not identical to those appearing at the top of

Table 6. Selected statistics from the multiple stepwise regression analysis with the selected dependent variable in the first column. The selected acoustic parameter is indicated at each step.

Variable	Step nr and variable	Multiple r	r^2	Beta weight
Hoarseness	1. PPQ	.598	.357	.598
Breathiness	1. NNEa	.734	.538	.734
Roughness	1. PPQ	.544	.296	.442
	2. APQ	.617	.381	.702
Gratings	1. PPQ	.393	.154	.393

Figure 2. PPQ and NNEa of the 50 child voices (upper and lower panels) versus mean perceptual rating of hoarseness. The line shows the best linear fit.

the perceptual rank order except for NNEa and *breathiness*, where the four voices rated as most breathy also appear at the top of the acoustic evaluation. However, the order of the voices is not identical. Regarding *hoarse*, *breathy* and *rough* compared to the PPQ measurement, the voice rated as second to most *hoarse* and *rough* received the highest PPQ value. The perceptual evaluation of *breathiness* ranked this voice tenth from the top.

Discussion

Despite rather high degrees of correlations, the scatter plots in Figures 3, 4, and 5 show a

considerable spread. Similar degrees of scatter have been found in corresponding analyses of adult voices (Askenfelt & Hammarberg, 1986). The scatter suggests an explanation why no correlation was found in our previous pilot study of six children's voices where relations between two perturbation measures and perceptual voice characteristics were analysed. Even though that study concerned running speech, it still seems difficult to predict the perceptual qualities of voices from acoustic analysis.

Against this background it seems appropriate to discuss the validity of these acoustic measures. It might be argued that a significant

Fig. 3. PPQ, NNEa and NNEb of the 50 child voices (upper, middle and lower panels) versus mean perceptual rating of breathiness. The line shows the best linear fit.

Fig. 4. PPQ and NNEa of the 50 child voices (upper and lower panels) versus mean perceptual rating of roughness. The line shows the best linear fit.

correlation with a perceptual characteristic is the criterion for validity of an acoustic measure. However, a perfect correlation between perceptual and acoustic voice characteristics is hardly to be expected. While an acoustic analysis reflects a set of well-defined aspects of the voice signal, perception would be highly influenced by the listener's previous experiences and internal standards. Therefore, perception is likely to depend on constellations of acoustic properties rather than on any one particular physiological aspect. Discarding all acoustic measures that fail to correlate with a perceptual parameter seems inappropriate. Physiological relevance would seem to be a better criterion for validity.

Still, some significant correlations were observed. All four acoustic measures were found to correlate with *breathiness*, with regard to PPQ, NNEa and NNEb at the highest level of

significance. In addition, PPQ and NNEa correlated with *hoarseness* and *roughness*. On the other hand, *instability* and *hyperfunction* did not correlate with any of the acoustic parameters. Regarding *instability* it may seem somewhat unexpected that no correlation was found with PPQ since fluctuations of fundamental frequency has generally been found to correlate with this perceptual parameter. However, it is possible that the fluctuations of fundamental frequency that occurred in our material were slow enough to be captured by the running F0 mean of the program and thus failed to affect the perturbation measures. Another explanation is that what our listeners perceived as *instability* was dependent on spectrum fluctuations that were not captured by our acoustic measures. With regard to *hyperfunction* it has

Fig. 5. Rank ordered ascending mean values for the four acoustic measures.

been pointed out that this perceptual variable is not necessarily associated with high jitter or harmonic-to-noise values (Klingholz & Martin, 1985). This result is not very surprising since *hyperfunction* mainly correlates with a low spectral level of the fundamental relative to the first formant. Thus, in order to identify *hyperfunctional* voices other acoustic measures than those used in the present study seem appropriate (Hammarberg, 1986; Hammarberg & Gauffin, 1995; Hillenbrand et al., 1994), viz., the amplitude of the first harmonic and the spectral tilt.

A recent study of adult voices showed that a perceptual voice evaluation was more reliable than an acoustic evaluation performed by three different automated systems (Rabinov et al., 1995). With increasing degree of voice deviation the automated analysis systems tended to break down whereas the listeners' reliability tended to improve. The authors proposed that acoustic measures may be more reliable and possibly preferred when discriminating between essentially normal voices. As none of the voices analysed in the present investigation were clearly pathological, the acoustic analysis could thus be expected to yield reasonably reliable results.

The correlations between the four acoustic parameters shown in Table 2 seemingly suggests that they at least partially reflect the same voice

feature. Indeed, all these parameters are affected by the periodicity of the fundamental frequency. Moreover, the PPQ and NNEa measures both yielded significant correlations with the same three perceptual parameters, thus suggesting an interrelation between these two measures. However, the correlations were not exactly the same. In the present study, PPQ showed high correlations with both APQ and NNEa. The relationship between jitter and harmonics-to-noise ratio has previously been observed by Yumoto and colleagues (1984) and by Wolfe et al. (1995). Yumoto et al. suggested that jitter contributes to the magnitude of the noise components in the harmonics-to-noise-ratio (HNR). On the other hand, it is also very likely that simultaneous combinations of these acoustic characteristics, as well as the perceptual, occurred in the voices analysed. This is exemplified by the results from the perceptual evaluation where the voice rated as most *hoarse* also had the highest rating on *breathiness* and *roughness*.

The rank ordered means for NNEa and NNEb did not show any marked discontinuities as could be observed in the distribution for PPQ and APQ. Still the correlation between particularly NNEa and *breathiness* was high. However, the rank ordered mean of *breathiness* has been shown to have a similar distribution as the NNE

measures (Sederholm et al., 1993). This may indicate that *breathiness* is a more acceptable voice trait in childrens voices, regarded as a less pathological parameter than *roughness* and *hoarseness*, where marked discontinuities in the rank ordered mean values could be observed (Sederholm et al., 1993).

The correlations differed slightly between boys' and girls' voices. For the girls', *breathiness* correlated only with PPQ and NNEa at the highest level of significance. The same applied to the boys' voices except that also NNEb showed a correlation. Although at a lower level of significance, the same difference occurred between girls and boys with respect to *hoarseness*. Regarding *roughness*, PPQ and NNEa correlated only for the boys' voices. Thus, the spectral noise measures seemed more relevant to perceived voice quality in boys than in girls. The boys' voices were perceptually evaluated as more *hoarse* and *rough* than the girls' voices. Only one girls' voice was perceptually rated to have a substantial degree of *roughness*. This may reflect a difference in vocal behaviour between the girls and the boys.

How representative of the normal phonatory habits are sustained vowels as compared with running speech? Analysing adult voices, Södersten & Hammarberg (1989) found that most perceptual parameters were more prominent in connected speech as compared to sustained vowels. Furthermore, with regard to test-retest reliability, reading a text has been found to yield more stable data than both running speech and sustained vowel phonation (Fitch, 1990). Sederholm (1996) recently carried out a perceptual study of the same 50 voices as were analysed in the present investigation. She found that expert listeners perceived the voices to be more unstable but less hoarse when sustaining a vowel as compared to running speech. Furthermore, in sustained vowels she found *roughness* to be the most important predictor of *hoarseness* while in running speech it was *hyperfunction*. It is possible that *roughness* is considered to be more pathological when present in a sustained vowel than in connected speech. Also, *instability* was almost absent in running speech but an important characteristic in sustained vowels. This is probably due to the more unfamiliar vocal task when sustaining a vowel. The fundamental frequency in phonation of a sustained vowel may also be relevant to consider. The children were instructed to phonate at a pitch level representative of normal speech. In some cases this turned out to be a difficult task, resulting in a fundamental frequency which appeared to be slightly above the habitual value.

Thus, running speech and sustaining a vowel seems to evoke a slightly different phonatory behaviour in children. Furthermore, it has been proposed that articulation and supraglottal constrictions for voiced continuants perturb the vocal fold vibrations (Bickley & Stevens, 1986; Stevens, 1991). Still, it seems preferable to perform the acoustic analysis on running speech in order to get a representative voice sample, especially in cases of essentially normal voices.

High frequency noise has been found to be an important component of voice deviation (Yanagihara, 1967). This is in accordance with our finding that *breathiness* correlated with all acoustic measures. Period perturbation quotient PPQ together with NNEa were the most effective acoustic parameters in detecting voice deviation, and NNEa reached the highest correlation with *breathiness*, $r=.734$. This suggests that these measurements, and in particular PPQ and NNEa, can be useful tools in voice evaluation. Kreiman et al. (1993) pointed out that the perceptual evaluation of voice quality is the standard against which all other measurements are compared and evaluated. On the other hand, acoustic measures may reveal essential aspects of vocal status, even if they cannot be readily perceived from a voice recording.

Conclusion

Our results revealed a reasonable correlation between the perceptual *hoarseness*, *breathiness* and *roughness* and the acoustic measures PPQ and NNEa. *Breathiness* correlated with PPQ, APQ, NNEa and NNEb. No correlation was found between *hyperfunction* and *instability* and any of these measures. APQ correlated only with *breathiness*. Based exclusively on the present acoustic measures voices with high degrees of *breathiness*, *hoarseness* and *roughness* would be identified while voices with *hyperfunction* and *instability* would not be detected.

Acknowledgement

The kind cooperation of all children is gratefully acknowledged as is the collaboration of our colleagues Jenny Iwarsson and Elisabeth Sederholm. We received valuable comments on the manuscript from Britta Hammarberg and Jan Dalkvist. This study was made possible by a grant from Axel and Margaret Ax:son Johnson's research foundation and Nämnden för Vårdvetenskap (NVV) (the Committee for the Health and Caring Sciences) at the Karolinska Institute.

References

- Arnold KS & Emanuel FW (1979). Spectral noise levels and roughness severity ratings for vowels produced by male children. *Journal of Speech and Hearing Research*, 22: 613-626.
- Askenfelt AG & Hammarberg B (1986). Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures. *Journal of Speech and Hearing Research*, 29: 50-64.
- Bickley C & Stevens KN (1986). Effects of vocal tract constriction on the glottal source: Experimental and modeling studies. *Journal of Phonetics*, 14: 373-382.
- De Krom G (1994). Consistency and reliability of voice quality ratings for different types of speech fragments. *Journal of Speech and Hearing Research*, 37: 985-1000.
- Eguchi S & Hirsh IJ (1969). Development of speech sounds in children. *Acta Otolaryngologica Supplementa* 257: 5-43.
- Fex B, Fex S, Shiromoto O & Hirano M (1994). Acoustic analysis of functional dysphonia: Before and after voice therapy (accent method). *Journal of Voice*, 8: 163-167.
- Fitch JL (1990). Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading and connected speech. *Journal of Speech and Hearing Research*, 55: 360-363.
- Hammarberg B (1986). *Perceptual and acoustic analysis of dysphonia*. Stockholm: Diss Dept. of Logopedics and Phoniatrics, Karolinska Institute, Stockholm.
- Hammarberg B. Personal communication.
- Hammarberg B & Gauffin J (1995). Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological aspects. In: Fujimura O & Hirano M, eds. *Vocal Fold Physiology. Voice quality control*, San Diego: Singular Publishing Group, 283-303.
- Hillenbrand J, Cleveland RA & Erickson RL (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* 37: 769-778.
- Kane M & Wellen CJ (1985). Acoustical measurements and clinical judgments of vocal quality in children with vocal nodules. *Folia Phoniatrica* 37: 53-57.
- Kasuya H, Ogawa S, Mashima K & Ebihara S (1986). Normalized noise energy as an acoustic measure to evaluate pathologic voice. *Journal of the Acoustic Society of America*, 80: 1329-1334.
- Kikuchi Y, Uchida S & Kasuya H (1986). An integrated voice analyzer for acoustic evaluation of pathological voice. *The Transactions of the IECE of Japan*, E 69; 10: 1057-1059.
- Klingholz F & Martin F (1985). Quantitative spectral evaluation of shimmer and jitter. *Journal of Speech and Hearing Research*, 28: 169-174.
- Koike Y (1973). Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonologica*, 7: 17-23.
- Kreiman J, Gerratt BR, Kempster GB, Erman A & Berke GS (1993). Perceptual evaluation of voice quality: Review, tutorial and a framework for future research. *Journal of Speech and Hearing Research*, 36: 21-40.
- McAllister A, Sederholm E, Sundberg J & Gramming P (1994). Relations between voice range profiles and physiological and perceptual voice characteristics in ten-year-old children. *Journal of Voice*, 8: 230-239.
- McAllister A, Sederholm E, Ternström S & Sundberg J (1996). Perturbation and hoarseness. A pilot study of six children's voices. *Journal of Voice*, 10/3: 252-261.
- Murry T & Doherty ET (1980). Selected acoustic characteristics of pathologic and normal speakers. *Journal of Speech and Hearing Research* 23: 361-369.
- Norusis MJ (1990). *SPSS/PC+ Statistics Base Manual, Statistics 5.0, and Advanced Statistics 5.0*. Chicago, SPSS Inc.
- Oates JM, Kirkby RJ (1980). An acoustical investigation of voice quality disorders in children with vocal nodules. *Australian Journal of Human Communication Disorders* 8;1.
- Pabon JPH (1991). Objective acoustic voice-quality parameters in the computer phonetogram. *Journal of Voice*, 5/3: 203-216.
- Rabinov CR, Kreiman J, Gerratt BR & Bielamowicz S (1995). Comparing reliability of perceptual ratings of roughness and acoustic measures of jitter. *Journal of Speech and Hearing Research* 38: 26-32.
- Sederholm E (1995). Prevalence of hoarseness in ten-year-old children. *Scand Journal of Logopedics and Phoniatrics*, 20: 165-173.
- Sederholm E (1996). Perceptual evaluation of connected speech and sustained vowels in children's voices. *Submitted for publication*.
- Sederholm E, McAllister A, Sundberg J & Dalkvist J (1993). Perceptual evaluation of hoarseness using continuous scales. *Scand Journal of Logopedics and Phoniatrics*, 18: 73-82.
- Stevens KN (1991). Vocal-fold vibrations for obstruent consonants. In: Gauffin J & Hammarberg B, eds. *Vocal Fold Physiology. Acoustic, Perceptual and Physiological Aspects of Voice Mechanisms*. San Diego: Singular Publishing Group, 29-36.
- Södersten M & Hammarberg B (1989). Perceptual analysis of voice quality in sustained vowels and connected speech. In: *Proceedings from the XX¹st Congress of the International Association of Logopedics and Phoniatrics*, Prague, II: 361-363.
- Wewers ME & Lowe NK (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health*, 13: 227-236.
- Wolfe V, Fitch J & Cornell R (1995). Acoustic prediction of severity in commonly occurring voice problems. *Journal of Speech and Hearing Research*, 38: 273-279.
- Yanagihara N (1967). Significance of harmonic changes and noise components in hoarseness.

- Journal of Speech and Hearing Research*, 10: 531-541.
- Yumoto E, Sasaki Y & Okamura H (1984). Harmonic-to-noise ratio and psychophysical measurement of the degree of hoarseness. *Journal of Speech and Hearing Research* 27: 2-6.