Bäckström T, Lehto L, Alku P, Vilkman E, Automatic pre-segmentation of running speech improves the robustness of several acoustic voice measures, Logopedics Phoniatrics Vocology, 2003; 28 (3): 101-108.

Automatic pre-segmentation of running speech improves the robustness of several acoustic voice measures

Tom Bäckström[1], Laura Lehto[1,2], Paavo Alku[1] and Erkki Vilkman[2,3]

From the [1]Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, P.O. Box 3000, FIN-02015 Hut, Finland, [2]Phoniatric Department, ENT Clinic, Helsinki University Central Hospital, P.O.B. 220, FIN-00029 Huch, Finland, and [3]University of Oulu, Department of Phoniatrics, FIN-90014 Oulun Yliopisto, Finland

Received 27 May 2002. Accepted 18 June 2003.

Logoped Phoniatr Vocol 2003; 28: 101–108

In order to study vocal loading, we developed a speech analysis environment for continuous speech. The objective was to build a robust system capable of handling large amounts of data while minimizing the amount of user-intervention required. The current version of the system can analyze up to five-minute recordings of speech at a time. Through a semiautomatic process it will classify a speech signal into segments of silence, voiced speech and unvoiced speech. Parameters extracted from the input signal include fundamental frequency, sound pressure level, alpha-ratio and speech segment information such as the ratio of speech to silence. This paper presents results from the performance evaluation of the system, which shows that the analysis environment is able to perform robust and consistent measurements of continuous speech.

Key words: continuous speech, speech analysis, vocal loading.

Correspondence: Tom Bäckström MSc, Helsinki University of Technology, Acoustics Lab, P. O. Box 3000, FIN-02015 Hut, Finland. Tel.: +358-9-4515843. Fax: +358-9-460224. E-mail: tom.backstrom@hut.fi

INTRODUCTION

Speech and voice professions have become commonplace in modern society (9, 18, 19). Extensive voice usage causes voice loading which, in turn, can result in occupational health problems (19). Previous research has been carried out on the effects of prolonged voice use on voice production (e.g., (10, 13)) but results on how loading is reflected in acoustical parameters are sparse. The most important finding to indicate vocal loading is the rise of fundamental frequency (F0). However, such concomitant issues as environment, speech task, speaker training, as well as the psychological state of the speaker, can cause discrepancies in the results and must be taken into account (12, 13, 16).

Most studies on speech have focused on stationary voice qualities whereas only a limited amount of work has been devoted to dynamic variations and the influence of a linguistic-phonetic frame including prosody (5). Concurrently, analysis of sustained vowels or fixed words is appealing for the technical simplicity of the analysis, but it is not clear how much this unnatural speech context affects the results (8, 14). Fast variations in running speech cause technical difficulties for the analysis task. For example, a vowel might not last more than 20 msec but still it should be classified as a vowel. This sets a lower limit to the granularity of parameters measured. Still, it is desirable to present some temporal averaging in order to minimize measurement errors due to the large signal variations.

In 1993, Kari Haataja had developed a speech analysis environment for the clinical speech analysis laboratory at Oulu University Hospital, Finland (6, 15). It is based on analog measurements of voice quality parameters which are transferred to a computer through a multi-channel A/D-card. The purpose of the current software was to update and replace this analysis environment with a modern digital one. The new environment implementation was to be based on state of the art signal processing techniques, which can now be implemented due to high-speed processors and large memory capacity.

The project was initiated by the largest Finnish telecommunications operator, Sonera, in an effort to gather data on the voices of their call-centre personnel. The objective was to improve the working environment of the call-centre personnel and to reduce the

ISSN 1401-5439

DOI: 10.1080/14015430310015237

number of sick-leave absences related to voice failures. It had been noted that the call-centre personnel had more sick leave than any other department and the company wished to analyze the reasons behind this.

The purpose of this paper is to evaluate with objective measurements the performance of the software. We will study the effects of user-induced variance in measurements, as well as the performance in different background noise conditions. A more extensive, technical description of the methods used in this project can be found in (3, 4).

SPEECH MATERIAL AND OPERATORS IN SYSTEM EVALUATION

Speech material

The voice material used was excerpts from customer service recordings at a call-centre. Throughout their working day, customer service employees answer the phone in an open-plan office. During phone calls, they discuss clients' problems related to their phone use. The entire pool of recorded voices consisted of 36 subjects (27 F, 9 M). A subset of these voices was used for this paper. Detailed analysis of all the voices is to appear in a separate study.

Speech was recorded on DAT tapes (sampling frequency 44.1 kHz) with Sony DAT TDC-D3 and TDC-D7 recorders, an AKG CK97-0 condenser microphone capsule and an AKG SE 300 B output module. The microphone was attached to the headset mouthpiece of the phone, close to the mouth (approximately 3 cm).

Operators

Performance of the system was evaluated by running a series of tests, where different operators used the system. Variance of the results was tested with three experiments. Firstly, intra-operator variance was measured by an experienced voice therapist with prior experience in the current software. She repeated a measurement five times on different days, using the same material (a five minute speech recording of a female subject).

Secondly, for inter-operator variance measurements, 13 subjects served as operators. Of these, six were speech professionals (speech-language therapists and similar) and seven were computer engineering professionals and students. Both groups included one operator with prior experience of the software, while the others were naïve users. Before the actual measurements, each naïve operator had a trial run supervised by one of the experienced operators. Each operator analyzed two recordings, one spoken by a female voice (age 22) and the other by a male voice (age 23), neither having a history of vocal disorders. Both recordings were chosen as average representatives of the voice material, in light of fundamental frequency and perceptual voice quality. The choice was made by an experienced speech-language therapist. The recording used in the trial run was the same female voice as in the actual measurement.

Thirdly, four operators analyzed four additional recordings, two male and two female voices. The recordings chosen were the voice samples with extreme (both lowest and highest) values of the average fundamental frequencies for each gender.

METHODS

Pattern recognition

The methodology essential in this work can be divided into three categories: Voice activity measures (also known as features), voice activity classification (also known as voice activity detection) and voice quality measures. The voice activity measures generate different features of the voice signal according to voice activity classification procedures. Each activity measure analyzes a different property of the signal, and the classification methods aim to combine these in an intelligent way in order to classify the signal into silence, voiced and unvoiced speech segments.

Once the classification has been carried out, the speech segments, both voiced and unvoiced, are analyzed using different voice quality measures. These quality measures are given to the user for further analyses.

The input signal is analyzed in 20 msec non-overlapping frames. For each frame, three parameters are extracted for the classification procedure: sound pressure level (SPL), fundamental frequency (F0) and a stationarity measure.

For the SPL measure, the user is given the option to calibrate the SPL or to apply the maximum-amplitude signal as a reference. Independently of the reference level, the SPL is thresholded to classify speech and silence segments.

The stationarity measure used is based on the Akaike information criterion, which is widely used in system identification (1, 17). This measure effectively estimates the amount of information in the autocorrelation, thus indicating the level of stationarity of the signal, as opposed to randomness. In our case, we defined the stationarity measure c as follows (1):

$$c = \frac{N\mathbf{r}^T\mathbf{r}}{r^2(0)}$$

where N is the number of elements in the autocorrelation vector $\mathbf{r}$. The Akaike information criterion is normally used for model-order specification. The criterion also contains model-order as a parameter, but in our case, the model is static and we therefore need only the simplified formula as above. The underlying assumption is that voiced speech is stationary whereas unvoiced speech is closer to white noise. The stationarity measure thus gives us the means to classify voiced and unvoiced speech.

The fundamental frequency was extracted with an autocorrelation-based algorithm. The algorithm locates the maximal correlation peak in a user-defined frequency range. If the F0 was outside the valid range, or a non-valid peak was located, the frame was classified as unvoiced even in those cases where the stationarity measure would indicate voiced speech.

Speech analysis

The signal features described above divided the input signal unambiguously into three classes: silence, voiced speech and unvoiced speech. However, the classifier sometimes generated erroneous results. To remove obvious classification errors, a post-classification procedure was applied to the classification vector. Speech and silence segment lengths were thresholded with constant parameters. These parameters are user defined, but typical values would be 250 msec for speech segment thresholding and 70 msec for silence thresholding. Speech segments are defined as the union of voiced and unvoiced speech. Speech segments shorter than the threshold were merged into silence segments, and silence segments similarly into unvoiced speech segments. This method ensures that voiced speech segments have valid F0 estimates while still removing most classification errors.

The segmented speech signal then provided all necessary data for the actual speech analysis. In addition to the measures described above (F0 and SPL), the alpha-ratio was also calculated. The alpha-ratio is defined as the ratio of energy in the signal above 1 kHz (up to the Nyquist frequency) and below 1 kHz (7). The alpha-ratio measure was applied to all frames in voiced speech segments thus eliminating the need to use long time average spectra (LTAS). Additionally, while each voiced speech frame is given equal weight, prosody will not introduce bias to the measure, as is the case with LTAS.

For all the measures, the program provides the following statistics: mean, standard deviation, median, histogram and time-domain profile. Optionally, each parameter can be filtered to remove 5% of the most extreme values (2.5% of highest and lowest values), in order to remove possible measurement errors (outliers).

A Signal to Noise Ratio (SNR) measure was also developed for evaluation of recording quality. In short, the SNR measure assumes that all silent segments are noise while speech segments are the desired signal. The ratios of the corresponding average energy levels are then compared to produce the SNR-value.

System environment

The software runs on a regular MS-Windows based off-the-shelf computer. The user-interface was created with National Instruments LabVIEW version 5.1 and mathematical programming in MATLAB version 6.0 R12 by MathWorks Inc.

The only hardware requirement, apart from the audio card, is the amount of RAM which should be at least 128 Mb. The program will run with less memory but in that case its speed is drastically reduced. The audio card used was a Turtle Beach Montego II+ that allows digital copying of DAT tapes and has excellent audio quality for analog recordings.

System usage

The analysis procedure consists of three stages for the operator: recording the speech, thresholding of signal energy and stationarity, and analysis of results. Recording usually consists of copying previously recorded DAT-tapes to the computer. Apart from setting the thresholding levels, the thresholding also includes visual and auditory identification and removal of non-speech sounds of high energy.

RESULTS

Intra-operator variance

In human-computer interaction, some variance will always appear due to the inexact behavior of the user. The current software contains manually tuned thresholds on energy and stationarity levels, which will undoubtedly present some user-induced intra-operator variance. The results were collected and compared to give an estimate of the measurement variance, as well as confidence intervals for the values measured. The measurement variances, and all other variance estimates in this paper, are normalized by N − 1 where N is the sequence length. This makes the estimate the best unbiased estimate. Furthermore, we present variances by means of standard deviations, since it is a more descriptive representation. The mean and standard deviation of the measurements are listed in Table 1.

The standard deviation of all parameters, with the exception of segment lengths, is approximately 1% of the mean or below. This indicates that these measure-

Table 1. *Intra-operator variance; one operator, five repeated measurements*

Parameter	Mean	Standard deviation	95%-interval	99%-interval
Total recording time (sec)	310.4	1.1	2.1	2.8
Speech time (sec)	111.4	1.4	2.7	3.5
SNR (dB)	16.00	0.12	0.24	0.31
Number of glottal oscillations	20560	210	400	530
SPL (dB)	84.5	0.2	0.4	0.5
F0 (Hz)	184.6	0.5	1.0	1.4
Alpha-ratio (dB)	17.77	0.05	0.09	0.12
Speech segment length (sec)	1.755	0.065	0.13	0.17
Voiced segment length (sec)	0.149	0.026	0.051	0.067
Unvoiced segment length (sec)	0.0702	0.0041	0.0080	0.0104
Silence segment length (sec)	2.59	0.06	0.12	0.16
Silence segment (< 2s) length (sec)	0.788	0.012	0.024	0.031
Silence segment (> 2s) length (sec)	6.84	0.11	0.22	0.28

ments are consistent and not affected by the user actions. The segment length measurements are technically more demanding and thus variance due to user decision will affect the results more easily. However, both speech time and speech segment length measures do not seem to deviate in subsequent measurements, indicating that the signal energy thresholding is an easier operation than the voiced/unvoiced segmentation. This is not surprising since the difference between voiced and unvoiced segments is difficult to distinguish even if it is done manually, let alone automatically.

It should be noted that the total recording time is not influenced by the actions of the operator, since the maximal recording time is a preset parameter of five minutes. Furthermore, the user-interface software, LabVIEW, regularly exceeds the preset recording length by a few seconds. This minor problem is a consequence of application synchronization difficulties due to the massive flux of data. However, since the preset recording time is so consistently exceeded (no cases of recording times below five minutes were observed), it was judged not to be a problem worth further action.

Since the 1% standard deviation seems to be a consistent approximation, it is suggested, as a rule of thumb, that the 95% confidence interval for measurements is 2% of the value measured.

Inter-operator variance

Variances of parameter values analyzed by different operators are shown in Table 2. In comparison to the intra-operator variance in Table 1, we can see that for the female voice the standard deviation is larger for the inter-operator measurements in three cases out of thirteen. However, the three cases that are larger in the inter-operator measurements, still fall well within the 95%-interval of measurements. It is therefore a plausible assumption that the intra-operator variance accounts for a larger part of the total variance than the inter-operator variance. Moreover, the three measurements that are larger in the inter-operator measurements, SPL, unvoiced segment length and silence segment (< 2 sec) length, are all among those measurements most sensitive to choices made by the operator.

The inter-operator variance for the male voice is, contrary to expectations, larger than that of the female voice. However, with the exception of fundamental frequency and the number of glottal oscillations, all measurements are insensitive to the common problems with the higher fundamental frequency of female voices. In addition, the difference in standard deviations between the male and the female voice is in most cases not significant.

Results of the third experiment, where four operators analyzed the female and male voices which had the extreme fundamental frequencies of the whole material, are listed in Table 3. The three most important measures for our purposes are F0, SPL and alpha-ratio, and their variances are therefore of special interest. Looking at Table 2 and Table 3, we can see that the ratio between the standard deviation and the mean (coefficient of variation) is, in all cases, below 2%, and, in 14 cases out of 18, below 1%. The largest value is found in Table 3 for the male with the highest F0 and for the F0 measure, where the coefficient of variation is 1.97%. It was to be expected that high-pitch voices are the most problematic cases, especially for the F0 measure. However, the worst case (standard deviation of 2%) still clearly indicates that the performance of the system is acceptable even for voices of extreme F0.

Table 2. *Inter-operator variance; a female and a male voice (with 13 operators), together with the trial run (for the 11 naïve operators)*

Parameter	Trial run (female voice)		Female voice		Male voice	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Total recording time (sec)	309.1	1.7	308.5	1.0	309.1	1.3
Speech time (sec)	112.5	1.1	112.1	0.8	75.0	2.4
SNR (dB)	15.98	0.15	16.09	0.10	17.07	0.42
Number of glottal oscillations	20760	170	20710	110	7420	220
SPL (dB)	84.07	0.46	84.11	0.28	81.26	0.65
F0 (Hz)	184.6	0.4	184.7	0.3	98.9	0.4
Alpha-ratio (dB)	17.78	0.05	17.80	0.04	16.74	0.07
Speech segment length (sec)	1.761	0.051	1.760	0.025	1.137	0.046
Voiced segment length (sec)	0.1819	0.0406	0.1845	0.0404	0.1877	0.0211
Unvoiced segment length (sec)	0.0619	0.0064	0.0637	0.0065	0.0659	0.0067
Silence segment length (sec)	2.56	0.05	2.58	0.04	3.34	0.09
Silence segment (< 2s) length (sec)	0.788	0.014	0.787	0.014	0.929	0.036
Silence segment (> 2s) length (sec)	6.74	0.13	6.68	0.11	6.43	0.23

Learning curve of operators

In the inter-operator variance measurements, each operator executed one trial run before the actual measurements, supervised by an experienced operator. Since the voice material in the trial run was the same as in the first actual measurement, it was possible to study how quickly the operator would learn how to use the software. Therefore, we compared all the resulting pairs of measurements for each naïve operator. We could not find any consistent trend in any of the measurement means. However, the inter-operator variance of each measure (the variance of measurements means over all operators) did decrease slightly in all cases, except for the average lengths of voiced and unvoiced segments, where no significant change in variance was found. The second, unsupervised, measurement (of the male voice) did not show a consistent trend in inter-operator variance with respect to the two earlier measurements. It is clear that this decrease in variance between the two first measurements is attributable to some learning process. Nevertheless, since the inter-operator variance for the male voice, in the majority of cases, was larger than that for the female voice (both in the trial run and the actual measurement), the learning process probably is associated with the specific voice, as opposed to learning of the software. In conclusion, learning the software is straightforward and does not seem to present any substantial variance to the results. Detailed results indicating operator learning are listed in Table 2.

Robustness of voiced/unvoiced classification

In addition to tests evaluating operator-induced variance, the second major task was to evaluate the performance of the stationarity measure. That is, to determine the SNR limit where a voiced sound cannot be distinguished from unvoiced sounds. Since the stationarity test is immune to changes in energy it can safely be assumed that unvoiced sounds and background noise are equivalent. With this objective, test sounds were created with increasing levels of noise added to concatenated words of /pa:p:a/ (Finnish for grandpa) pronounced by one female and one male subject (the same speakers as before). There was a silence segment in between the consecutive /pa:p:a/ words such that the length of a sample was one second. The SNR range was from 20 dB down to 0 dB in 2 dB steps. Gain was adjusted so that the overall SPL was constant.

Two noise types were used: white noise and ambient noise. The ambient noise was synthesized so as to mimic the real background noise present in the recordings. In order to model the background recording noise, we searched a non-speech segment (of 10

Table 3. *Inter-operator variance; voices with highest and lowest F0 for both female and male voices, with 4 operators*

Parameter	Female (Highest F0)		Female (Lowest F0)		Male (Highest F0)		Male (Lowest F0)	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Total recording time (sec)	308.1	1.0	310.1	1.9	309.1	1.0	310.6	1.2
Speech time (sec)	78.8	2.3	79.9	1.4	48.3	2.7	71.5	1.7
SNR (dB)	12.17	0.19	15.87	0.31	13.53	0.66	19.78	0.40
Number of glottal oscillations	15840	420	13930	190	5450	270	6740	220
SPL (dB)	78.5	0.6	82.1	0.7	80.9	1.1	86.6	0.9
F0 (Hz)	201.1	0.6	174.3	0.8	113.0	2.2	94.2	1.0
Alpha-ratio (dB)	17.23	0.05	19.07	0.15	18.12	0.24	17.21	0.12
Speech segment length (sec)	1.350	0.058	1.172	0.054	0.878	0.046	1.230	0.035
Voiced segment length (sec)	0.1580	0.0345	0.1782	0.0169	0.1655	0.0521	0.2145	0.0129
Unvoiced segment length (sec)	0.0788	0.0021	0.0725	0.0024	0.0738	0.0090	0.0773	0.0034
Silence segment length (sec)	2.87	0.05	2.86	0.09	3.75	0.25	3.80	0.07
Silence segment (< 2s) length (sec)	0.855	0.020	1.063	0.056	0.989	0.059	1.310	0.042
Silence segment (> 2s) length (sec)	5.57	0.13	5.13	0.17	6.71	0.42	5.14	0.13

seconds) from the original recordings made in the open-plan office. A linear predictive model of order 6 was computed for this segment (11). Model order was chosen large enough to grasp the spectral envelope of noise but small enough not to model (possible) resonances. Finally, the model generated was excited with white random noise to produce a noise signal with spectral envelope properties similar to those of the background noise in the original recordings.

The main result with both white and ambient noise was that both speech and voiced segments were easily detected from the noise-corrupted vowels in the whole SNR range, and for both subjects. For the female voice, speech segments and voiced segments were all classified correctly. For the male voice, speech segments were correctly classified with SNR values from 20 dB to 4 dB. For SNR values of 2 dB and 0 dB, the speech segments identified by the classifier were slightly longer than expected. However, the voiced/unvoiced classification was successful in all cases. It is unlikely that any recording would have a worse SNR than 0 dB. We can therefore safely conclude that voiced/unvoiced segmentation is not a problem with respect to SNR level.

Robustness of parameter computation

The third objective was to evaluate how changes in the recording conditions affect the results. For this purpose, 10-second speech excerpts of the recordings of both subjects were analyzed in the following manner. Eleven different levels of noise (SNR level ranged from 30 dB down to 0 dB with 3 dB intervals) were added to the speech signal and the results were compared. The noise signals were the same as those described in the previous section.

For signals corrupted with white noise, all values, except for the alpha-ratio and the SNR estimate, were found to be approximately constant over the whole SNR range. If the measurements had any trend it was masked by the inter-measurement variance. A slight increase in variance could have been expected but it is difficult to distinguish such a change in measurements without a large number of laborious repetitions.

With the decreasing SNR a clear increase in the alpha-ratio was evident as was to be expected. White noise added to the speech signal has more energy at high frequencies than the original speech signal. An increase in the level of the noise component will thus increase the relative amount of high frequencies, thereby increasing the alpha-ratio.

In the measurements, the SNR estimate decreased proportionally to the increase of the noise component. However, the range of the SNR estimate was not quite as large as the ratio of the energies of the original speech signal and the noise component. This is due to the fact that the clean speech signal always contained some noise and the SNR estimate is thus always biased.

For the signal corrupted with ambient noise, the results were similar to those obtained using white noise. The fundamental frequency estimates were within 1 Hz up to the SNR of 6 dB for both the female and male voice. Alpha-ratio was approximately constant up to an SNR of 9 dB where it started to rise, similarly to white noise. With the increase of noise, weak segments of speech disappear slowly into the noise. Speech segment lengths were therefore observed to shrink with decreasing SNR. Furthermore, voiced segments sometimes shrunk and scattered, since weak voiced segments had less energy than the threshold. However, this effect often did not split the speech segment, since energy dips of short duration (such as voiceless plosives) were ignored. Nevertheless, one should not be too concerned about these problems since they appeared only on SNR levels below 9 dB.

No significant differences in the results between the male and female subject were found in these measurements.

It should be noted that the ambient noise signal was concentrated heavily on the lower end of the spectrum. This could degrade the accuracy of energy thresholding of the signal as well as estimates of the fundamental frequency. However, in the current case, this did not present a problem. Should this become a problem in some other setting, it could be easily solved by high-pass filtering the signal with a cut-off of, e.g., 60 Hz. An option for filtering is built into the software, but as explained above, it was thus not used in the current experiments.

CONCLUSIONS

In this paper, we have evaluated the performance of an analysis environment for studying effects of vocal loading from continuous speech. The performance was evaluated by means of repeated measurements of the same signal, vowels in noise, and speech samples in noise. The results show that the analysis software performs consistently, that is, user actions do not significantly alter the results. It was shown that the 95% confidence interval for measurements is 2% of their average magnitudes. Furthermore, recording conditions, that is, the SNR level of measurements, do not degrade the quality of measurements. Only the alpha-ratio increases with a decreased SNR. However, this is not a design flaw of the software but rather a property of the measure.

This software was developed for vocal loading measurements. As expansions to the current software, we have considered inclusion of automatic inverse filtering of the voice source in the system, using, for example, the method presented in (2). Performing inverse filtering on voiced speech segment is supported by the classification algorithm implemented in the system. Moreover, this classification algorithm is not restricted to vocal loading studies alone, but can be applied in any task requiring voice/unvoiced classification of continuous speech.

In conclusion, we have described and evaluated in this paper the performance of an analysis environment for continuous speech. This analysis shows that the environment is able to perform robust and consistent measurements of continuous speech.

ACKNOWLEDGEMENTS

The project was funded by the Finnish Work Environment Fund and the Academy of Finland (Project Number 200949). It was carried out in collaboration among Sonera, Oulu University Hospital, Helsinki University Central Hospital and the Laboratory of Acoustics and Audio Signal Processing, at the Helsinki University of Technology.

REFERENCES

1. Akaike H. A New Look at the Statistical Model Identification. IEEE Trans Auto Control 1974; 19: 716–23.
2. Alku P. Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering. Speech Commun 1992; 11: 109–18.
3. Bäckström T, Alku P, Vilkman E. An analysis environment for studying effects of vocal loading from continuous speech. In Proceedings of the 5th International Conference on Advances in Quant. Voice and Speech Research, Groningen, the Netherlands, April 2001.
4. Bäckström T. Development of an analysis environment for clinical analysis of continuous speech [Master's Thesis]. Helsinki University of Technology, Helsinki, Finland; 2001. Available from: http://www.acoustics.hut.fi/publications/files/theses/backstrom_mst.pdf
5. Fant G. The voice source in connected speech. Speech Commun 1997; 22: 125–39.
6. Haataja K. Kliinisen puheentutkimuslaboratorion puheenanalysointi-ympäristö (A speech analysis environment for a clinical speech analysis laboratory) [Master's Thesis]. Oulu University Hospital, Oulu, Finland; 1993.
7. Kitzing P. LTAS criteria pertinent to the measurement of voice quality. J Phon 1986; 14: 477–82.
8. Klingholz F. Acoustic recognition of voice disorders: A comparative study of running speech versus sustained vowels. J Acoust Soc Am 1990; 87: 2218–24.
9. Laukkanen A-M. On Speaking Voice Exercises. Academic dissertation. Acta Universitatis Tamperensis, ser A, Vol. 445, Tampere, University of Tampere; 1995.
10. Lauri ER, Alku P, Vilkman E, Sala E, Sihvo M. Effects of Prolonged Oral Reading on Time-Based Glottal Flow

Waveform Parameters with Special Reference to Gender Differences. Folia Phoniatr Logop 1996; 49: 234–46.

11. Makhoul J. Linear prediction: A tutorial review. Proceedings IEEE April 1975;63:561–80.
12. Ohlsson A-C, Löfqvist A. Work-day Effects on Vocal Behaviour in Switchboard Operators and Speech Therapists. Scand J Logop Phoniatr 1987; 12: 70–9.
13. Pausewang GM, Andrews ML, Schmidt CP. Effects of Prolonged Loud Reading on Selected Measures of Vocal Function in Trained and Untrained Singers. J Voice 1991; 5: 158–67.
14. Qi Y, Hillman RE, Milstein C. The estimation of signal-to-noise ratio in continuous speech for disordered voices. J Acoust Soc Am 1990; 105: 2532–5.
15. Rantala L, Haataja K, Vilkman E, Körkkö P. Practical arrangements and methods in the field examination and speaking style analysis of professional voice users. Scand J Logop Phoniatr 1994; 19: 43–54.
16. Scherer RC, Titze I, Raphael BN, Wood RP, Raming LA, Blager FB. Vocal Fatigue in a Trained and an Untrained Voice User. In: Baer T, Sasaki C, Harris KS (editors). Laryngeal Function in Phonation and Respiration. Boston: College-Hill; 1987. p. 533–54.
17. Söderström T, Stoica P. System identification. Cambridge UK: Prentice Hall International, University Press; 1989. p. 423–6.
18. Titze I, Lemke J, Montequin D. Populations in the US Workforce Who Rely on Voice as a Primary Tool of Trade. NCVS Status and Progress Report 1996; 10: 127–32.
19. Vilkman E. Occupational Risk Factors and Voice Disorders. Logoped Phoniatr Vocol 1996; 21: 137–41.