The self-to-other ratio applied as a phonation detector for voice accumulation

Svante Granqvist

Abstract

A new method for phonation detection is presented. The method utilises two microphones attached near the subject's ears. In short, phonation is assumed to occur when the signals appear mainly in-phase and at equal amplitude. Several signal processing steps are added in order to improve the phonation detection, and finally the original signal is sorted in separate channels corresponding to the phonated and non-phonated instances. The method is tested in a laboratory setting to demonstrate the need for some of the stages of the signal processing and to examine the processing speed. The resulting sound file allows for measurement of phonation time, speaking time and fundamental frequency of the subject and sound pressure level of the subject's voice and the environmental sounds separately. The present implementation gives great freedom for adjustment of analysis parameters, since the microphone signals are recorded on DAT tape and the processing is performed off-line on a PC. In future versions, a voice accumulator based on this principle could be designed in order to shorten analysis time and thus make the method more appropriate for clinical use.

Introduction

In clinical work with voice patients, one major concern is the patient's vocal habits. It is known that persons that have occupations that rely on frequent use of the voice are overrepresented in voice clinics (Fritzell, 1996). However, it might be difficult to acquire appropriate information about such habits in a voice clinic with its typically quiet environment. A more realistic approximation of a working environment can be simulated by subjecting the patient to noise through headphones (e.g. Neils & Yairi, 1987) or even loudspeakers (Ternström et al., 2002). Such artificial environment, however, still may lack important aspects of conditions in the patient's every-day work that are relevant to vocal habits. The real environment may indeed be the crucial factor for eliciting a patients' typical vocal behaviour; for example, it is common to raise fundamental frequency (F_0) when speaking to children (van de Weijer, 2002).

The potential importance of environmental factors call for other methods to record voice habits than studio recordings in the clinic. So far, the only reliable way to acquire data valid for the normal working environment is to perform *voice accumulation* at the workplace. Such voice accumulation has been performed by means of custom-built devices, *voice accumulators*. These record the signal from a contact microphone or electroglottograph (EGG) applied to the subject's neck (Kitzing, 1979). In some cases, these recordings are complemented by a microphone that captures the airborne voice signal (Airo et al., 2000; Buekers et al., 1995).

Typically, such devices extract various features of the signal(s) and store in terms of statistical data about the patient's voicing. The output is thus extracts of predetermined parameters derived from analysis with predetermined settings. While this may be suitable in clinical practise, there is no opportunity to change settings or add analysis parameters. Also, certain analysis errors might be hard to discover. To avoid such problems, off-line analysis of continuously recorded voice signals is needed. Such a procedure could be particularly valuable during method development and research.

A typical voice parameter of interest is *phonation time*, i.e. the time that the vocal folds oscillate, usually expressed as a percentage of total recorded time. The *speaking time* is another relevant parameter. It is identical with phonation time except that unvoiced sounds and short mid-sentence pauses are included (Airo et al., 2000; Ohlsson et al., 1989; Szabo et al., 2001). Also F_0, and *sound pressure level* (SPL) are

frequently logged. From these parameters, different vocal dose measures have been derived (Titze et al., accepted for publ.)

There are, however, some difficulties involved with these methods for voice accumulation. A contact microphone must be firmly applied to the neck, and subcutaneous fat as well as beard and loose skin have been found to cause problems (Szabo et al., 2001). The signal amplitude from the contact microphone is not directly proportional to the SPL of the voice, neither does it accurately reflect vocal effort, which rather would be related to subglottal pressure and glottal adduction. Similar restrictions apply to electroglottography. Nevertheless, there have been attempts to estimate SPL by means of a contact microphone; Masuda et al. (1993) used a contact microphone signal to derive a rather coarse classification of sound levels, by dividing the amplitude from the contact microphone into four ranges. Popolo et al. (2002) used an accelerometer and a calibration procedure to derive SPL. The 95% confidence interval in this calibration was rather large; slightly more than 10 dB. This would make the method best suited for estimation of the long-time average of the sound pressure.

Thus, a microphone collecting the airborne voice signal appears desirable if the SPL of the voice is to be accurately measured. On the other hand, such a microphone will also pick up environmental sound. In some cases, the effect of such sound can be neglected (Buekers et al., 1995; Rantala et al., 1998) particularly if the microphone is placed near the subject's mouth and if the environmental SPL is low. However, there mostly will be a need for detecting the instances of phonation so as to differentiate the instances of the patient's own voicing from the instances containing environmental sound only. If such phonation detection is performed, also the SPL of background sounds can be estimated using the same microphone during non-phonated timeslots. Watanabe (1987) detected phonation by simultaneously recording a contact microphone signal, thus entailing the problems with the contact microphone signal quality mentioned. Airo et al. (2000) used two microphones placed at different distances from the mouth; if the level difference at the two microphones was above a threshold, the sound was assumed to originate mainly from subject.

Here an alternate method is presented that automatically detects phonation without the use of a contact microphone or EGG. The method uses two microphones for the airborne voice signal, applied near the subject's ears, at equal distance from the mouth. The basis for phonation detection is the fact that, under these conditions, signals from the speaker's own voice appear at equal amplitude and in phase at the microphones, while environmental signals usually do not. The method has been tested previously by Södersten et al. (2002) and Szabo et al., accepted for publ) and the aim of this paper is to provide a more detailed technical description of the signal processing.

Method

In the following, the method implemented in the custom made Aura computer program (Figure 1) is described. Whenever typical values for the parameters are given, they have been empirically determined as reasonable, but do not represent optimised values in the strict sense. The abbreviated names of the signals are collected in the appendix.

Detection of phonation

The method is based on estimation of the self-to-other ratio (SOR) as described by Ternström (1994). This method requires two omnidirectional microphones with identical frequency response and sensitivity, which are attached symmetrically near the subject's ears (Figure 2). Due to the symmetrical placement of the microphones, sound originating from the mouth of the subject appear at approximately equal magnitude and phase at the two microphones. On the other hand, ambient sounds most likely appear in different phase at the two microphones unless the origin is straight in front of or behind the subject, and within the reverberation radius of the room. From the two microphone audio (M_A) signals a difference channel Δ_A and a sum channel Σ_A is formed (Figure 3). In the difference channel, sounds originating from the subject are cancelled due to their in-phase relation at the microphones. In the sum channel, the subject's voice gains 6 dB as compared to the signal at each microphone, also due to the in-phase relation of the signals. For ambient sounds, which appear at the two microphones with low correlation, the corresponding gain is approximately 3 dB, both in the sum and difference channels. From the sum and difference channels, a SOR level was derived as the level difference between the two. Thus, in case of no phonation, the SOR level is typically

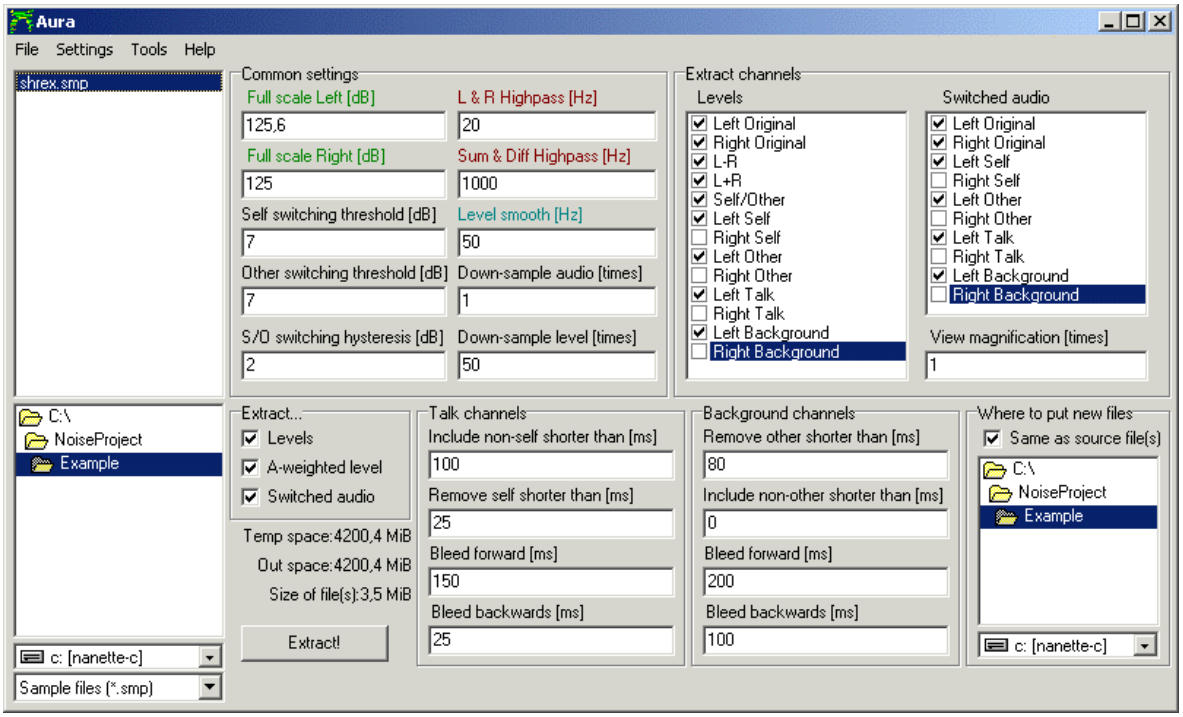

Figure 1. The user interface of the Aura computer program

0 dB, regardless of the ambient sound level. In the case where the subject phonates, the SOR will rise. However, the level of the phonation must be similar to or higher than that of the ambient sound in order to achieve a noticeable rise of the level. Luckily, subjects will in most cases automatically adapt their speaking level so that the sound level at their ears from the own voice is high, as compared to ambient sounds. Thus, a high SOR level will normally be strongly correlated to the instances when the subject speaks, thereby providing a potential basis for a phonation detector.

Figure 2. Placement of the ear microphones.

High-pass filtering

The method outlined above is based on the assumption that ambient sounds are uncorrelated at the two microphones. This is, however, not true for low frequency sounds, which appear in-phase at the two microphones, regardless of origin. Thus, in the presence of low-frequency ambient sounds only, there would be a high SOR even in the absence of phonation. To some extent this artefact was compensated for by applying a second-order high-pass filter to the M_A signals before extracting the SOR. This implies that more weight is attributed to the differentiating high frequency range than to the non-differentiating low frequency range. Taking into account the fact that counter phase occurs at a distance of 0.5 wavelength, and that a typical acoustic distance between the microphones is 20 cm, approximately, which corresponds to a half wavelength at f=860 Hz, the HP filter frequency could be somewhere near this frequency. In other words, the best cut-off frequency of the high-pass filter can be expected to lie in the

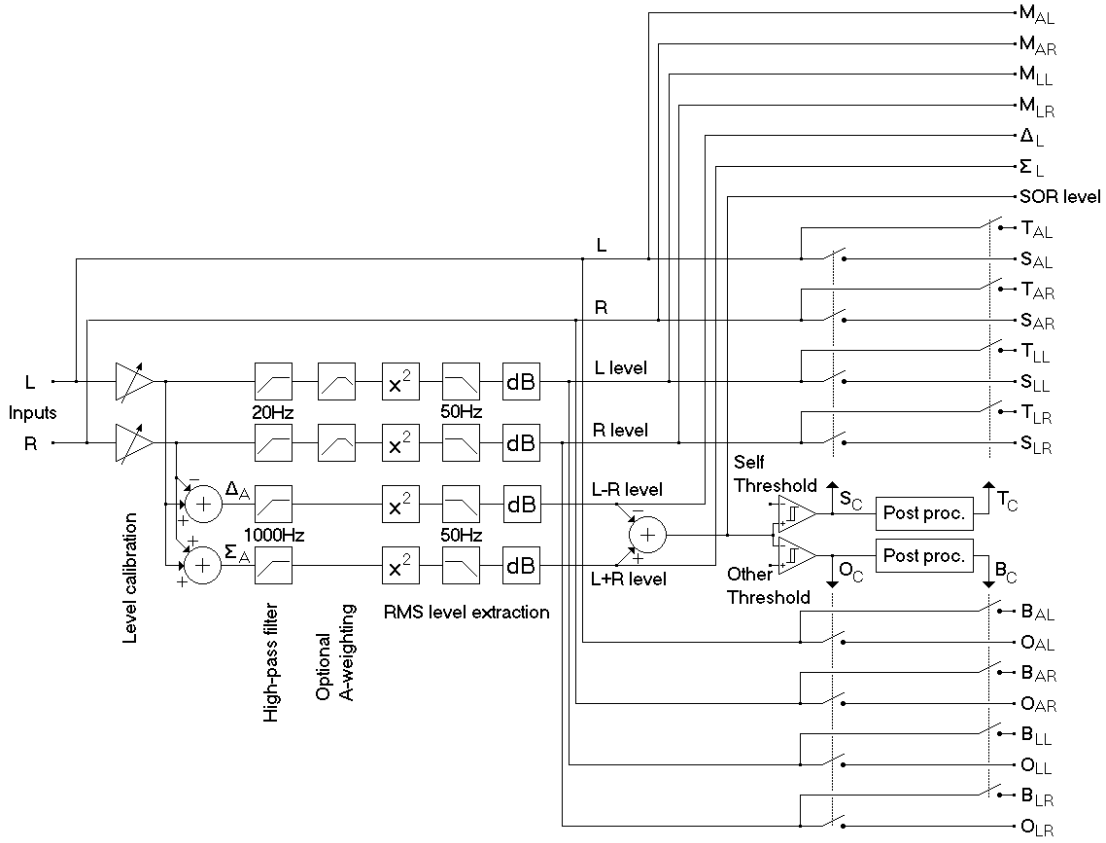

Figure 3. Block scheme of the signal processing performed by the Aura computer program.

range 500 – 1000 Hz. In the applications presented here, a Q-value of 1.53 was used for these filters.

Sorting

The SOR can thus be utilised as a criterion for phonation detection. Assuming that the SOR is correlated to the subject's phonation, it can be used to sort the M_A signals into separate channels, reflecting the presence or absence of phonation. Likewise, their two associated level signals M_L are sorted into separate channels. Ideally, these signals could then be used for separate measurement of voice and background sound. In order to perform such sorting, control signals need to be derived from the SOR.

Criteria for sorting

Self and Other signals were obtained by applying a threshold (typically 5-8 dB) to the SOR, and when the SOR exceeded this threshold, the control signal for Self (S_C) was considered as active. A complementary signal, the control signal for Other (O_C), was derived such that another threshold (typically 5-8 dB) was applied to the SOR and when the SOR was below this threshold, the O_C was considered as active. The two thresholds had hystereses (typically 2 dB) in order to reduce flickering between the active/inactive states. The state of S_C and O_C thereafter controlled the sorting. An active S_C signal implied that M_A was copied to the Self audio (S_A) channel, and that the M_L was copied to the Self level (S_L) channel. An active O_C signal caused M_A and M_L to be copied to Other audio (O_A) and Other level (O_L), respectively.

Post processing

Practical experience of the above procedure showed that the S_A channel typically lacked unvoiced consonants. In some cases, such a lack is desirable, for example for the purpose of estimating phonation time or fundamental frequency (F_0). In other cases, when an estimate of the speaking time is desired, these sounds should be included.

Also, some of the voicing from the subject could be heard in the O_A channel. A plausible reason for this appears to be the reverberation in the room, which causes the own voice to be reflected back to the microphones, appearing to be ambient sound. However, if the Other channel is to be utilised for measurement of background noise level, such leakage is undesirable. For these reasons, two additional control signals were derived by post-processing of S_C and O_C; the Talk control (T_C) and Background control (B_C) signals. For the T_C signal this post-processing was performed in the following steps

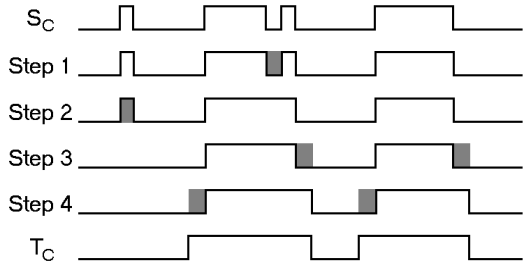

Figure 4. Steps for the post-processing of the Self channel control signal (S_C), resulting in the Talk channel control signal (T_C).

Start with a copy of the S_C signal
1. Include short inactive instances (typically < 100 ms)
2. Exclude short active instances (typically < 25 ms)
3. Move deactivation instances forward (typically by 150 ms)
4. Move activation instances backward (typically by 25 ms)

Typically these steps have the following effects. Step 1 includes segments containing unvoiced consonants. Step 2 removes intermittent ambient sounds. Step 3 includes the "tail" of the phonation and some reverberation and step 4, finally, adds some time at the beginning to ensure that the onset of phonation is included. The resulting audio channels T_A typically include also consonants and appear less "chopped".

For the B_C signal this post-processing was performed in the following steps

Start with a copy of the O_C signal
1. Exclude short active instances (typically < 80 ms)
2. Include short inactive instances (typically not used)
3. Move deactivation instances forward (typically by 200 ms)
4. Move activation instances backward (typically by 100 ms)

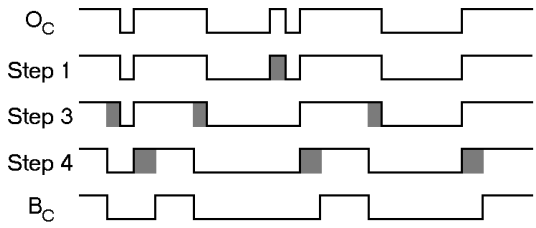

Figure 5. Steps for the post-processing of the Other channel control signal (O_C), resulting in the Background channel control signal (B_C). Here, as in most cases, step 2 is excluded.

Typically these steps have the following effects. Step 1 removes mid-sentence consonants. Step 2 is included in the computer program for symmetry, but is typically not used. Step 3 removes the reverberant part of voicing, and step 4 assures that the onset of voicing is properly excluded. The resulting audio channels B_A typically contain less sound originating from the subject.

Minor details

The Aura program also allows for the following adjustments:

- Level calibration of M_A.
- High-pass filtering of M_A; this feature is useful for removing low-frequency noise and DC drift from the SPL measurements. Filters are second order, Q-value 0.7071, typical cut-off frequency: 20 Hz.
- Optional A-weighting of levels.
- Smoothing filter used for level extraction. These filters consist of four cascaded first order links. Using first order links ensures that the step response has no ripple. Such ripple would be problematic for the following log conversion to decibels. Typical cut-off frequency: 50 Hz.

- Down-sampling of audio (typically not used).
- Down-sampling of level curves (typically 16 to 160 times).

Experiments

Södersten and collaborators (2002) tested the method under realistic conditions in a study of pre-school teachers' voice use during work. Here, some more specific aspects of the method are examined by means of three experiments. The first concerned the effect of ambient sound at low frequencies, which occurs in-phase in the two microphones. The second was a practical test of the program's automatic switching between two speakers during a conversation. The third concerned the processing speed of the Aura computer program. For these experiments, no DAT recorder was used; the microphone signals were recorded directly on a computer using a Sound Blaster Live! soundcard at a sampling rate of 48 kHz. The microphones were omnidirectional electret microphones (TCM 110).

S/O ratio spectrum for ambient white noise

Two loudspeakers were placed in a standard laboratory environment and fed with white noise. A subject was positioned in the diffuse field from these loudspeakers. The subject rotated slowly 360 degrees in order to expose the two microphones to the same average sound field. Long-time average spectra (LTAS) were used to analyse the spectral properties of the Σ_A and Δ_A channels, averaged over the entire rotation. The level difference between the two spectra exhibited a rise at low frequencies (Figure 6), even though the subject did not phonate during this experiment. At frequencies above 500 Hz, however, the level difference is near 0 dB, as predicted by theory. This explains the improved switching results obtained when the high-pass filter was introduced.

Conversation

A recording was made of a male and a female subject alternately reading a text. The sound was picked up by the microphones mounted at the ears of the female subject. White noise, filtered by a first order (-6 dB/octave) lowpass filter at 1 kHz, was presented over loudspeakers at an SPL of approximately 76 dB at the microphones. The waveforms and level curves shown in Figure 7 illustrate the results of the different steps in the signal processing. The two uppermost panels (A and B) represent the audio recording of the left and right channels together with their levels. Panel C shows the sum and difference levels and the level of the SOR, which shows high levels

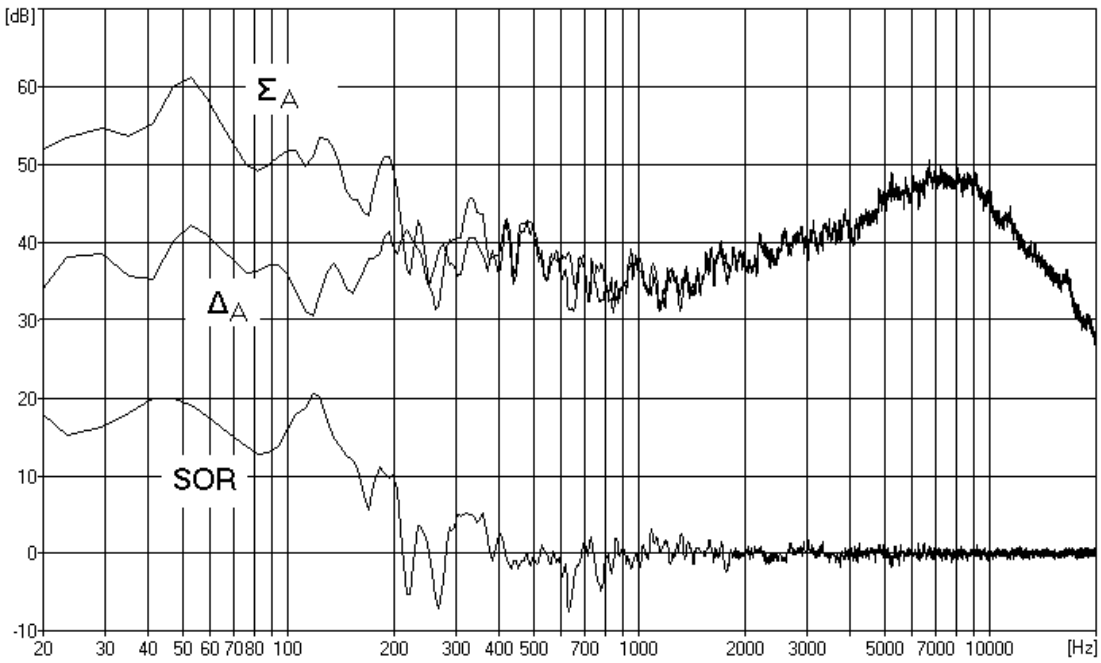

Figure 6. LTAS of sum and difference channels, and the SOR. Even though the subject was silent during this recording, the SOR exhibits a rise at low frequencies. This was the reason for high-pass filtering the sum and difference channels.

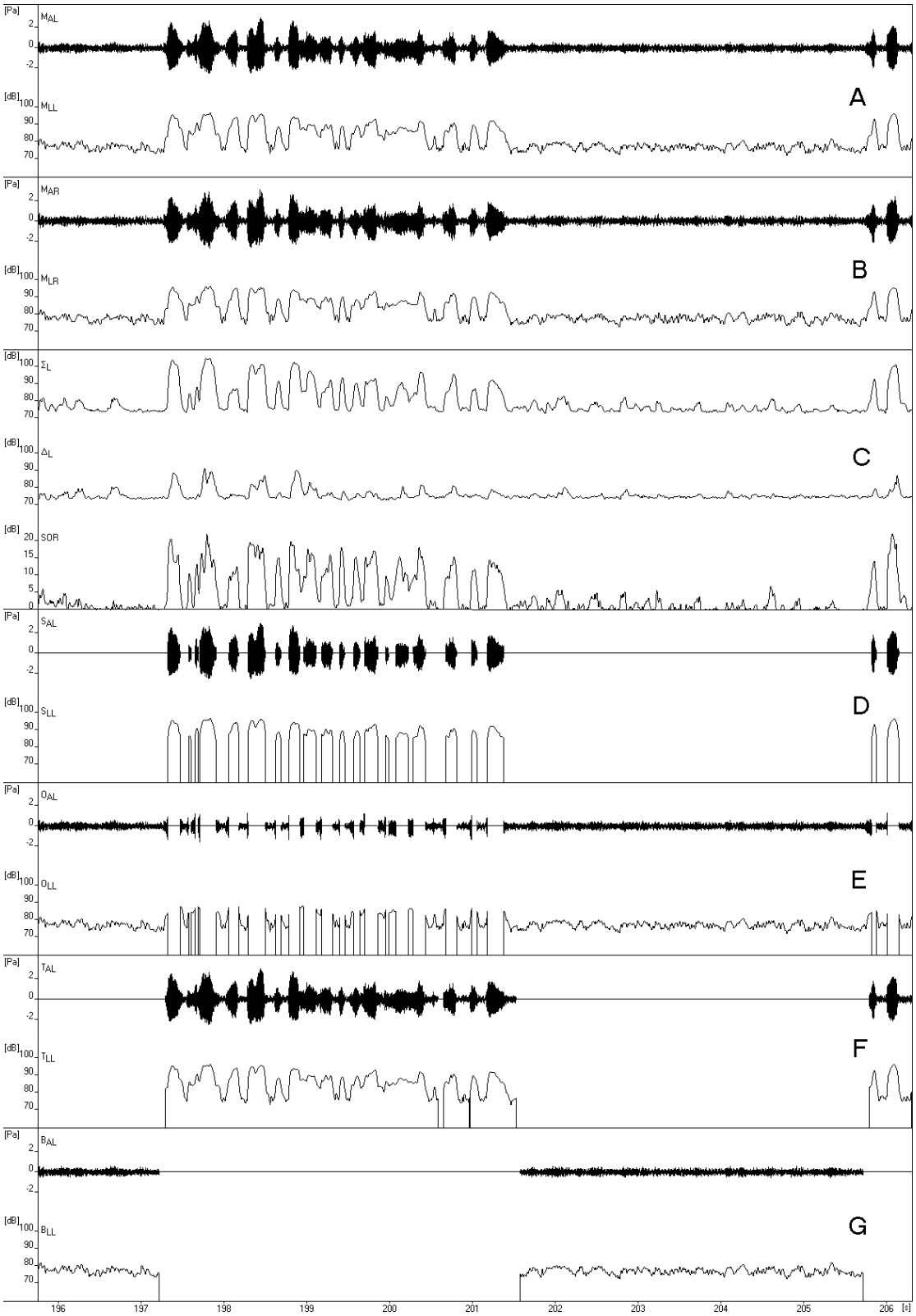

Figure 7. Example of signals from the Aura computer program. The subject is speaking between t=197.3 s and t=201.5 s. During this period the SOR signal (panel C) frequently reaches a high level. This fact is utilised to derive the Self, Other, Talk, and Background signals (panels D through G).

during the female speaker's utterances. Panel D shows the Self channels which are excerpts from the curves in panel A during instances of high SOR. Complementarily, panel E shows the Other channels which are excerpts from the curves in panel A during instances of low SOR and thus represents the instances pertaining to the loudspeaker noise plus the male subject's speech, which is mostly hidden under the loud noise. Panels E and F, finally, show the post-processed versions of the C and D curves, respectively. Channels corresponding to panels D, E, F and G but derived from the right ear microphone are also available, though not shown in the figure.

Processing speed

The processing speed of the Aura computer program was examined by processing a file of about 7 minutes duration on a 1200 MHz AMD Athlon computer running Windows ME. The Aura program allows extraction of audio channels and their corresponding level channels. The level channels can be both with and without A-weighting. In addition, sum, difference and SOR levels can be extracted, giving a total of 33 possible channels. If all these channels were extracted, the time needed was approximately 3.2 times the duration of the original file. However, the time was strongly dependent on the number of channels extracted. If only the S_{AL}, S_{LL}, B_{AL} and B_{LL} channels were extracted, the processing time was reduced to 0.4 times the duration of the original file.

Discussion and Conclusions

Although the method presented above seems to possess important advantages over currently available alternatives, it still has limitations that should be observed. First, a high SOR is obtained not only from the subject's own voicing, but also from sounds originating from all locations in the subject's mid-sagittal plane (straight behind, above etc the subject). In addition such sounds must originate within the reverberation radius of the room, and appear at the ears at a level close to or louder than the ambient sound. However, as all these conditions are rarely met simultaneously, this problem would be of minor concern.

Second, the symmetrical placement of the microphones is crucial. The precision required is determined by the wave-length, λ, of the sounds originating from the subject. When difference in microphone-to-mouth distance corresponds to $\lambda/4$, the SOR level will be 0 dB in the absence of ambient sound. The sounds from the subject mainly originate either from vocal fold oscillation, typically in the range 100-4000 with major frequency components between 500 and 1000 Hz, or from fricative sounds with maximum amplitude in the range 500-8000 Hz. At 8000 Hz $\lambda/4$ is about 1 cm and at 500 and 1000 Hz $\lambda/4$ is between 16 and 8 cm. Thus, symmetric placement of the microphones is important, the required precision lying in the range of centimetres.

Third, there might be occasions when the acoustic environment leads to asymmetrical transportation of the subject's voice to the microphones. A typical example would be telephone conversation, where the subject shields one of the microphones with the hand and telephone. However, apart from telephone calls, significant asymmetry appears to be unusual.

Fourth, the high-pass filtering required to reduce the effect of in-phase ambient sounds will also attenuate the lower partials of the subject's voice. This can be problematic if the subject has a voice that is dominated by the lower partials, such as in hypofunctional breathiness. In such cases, the voiced segments may incorrectly be detected as not being phonated. On the other hand, soft phonation would rarely be used in the presence of loud ambient sound.

A SOR of 0 dB reflects silence of the subject. This may appear somewhat counterintuitive since the true SOR level in this case is $-\infty$ dB, but it reflects the fact that the estimation method performs poorly at low SORs. This, however, is of little concern for the detection of phonation, since the detection criterion only depends on whether the SOR is above or below the threshold.

The method has been found to yield reliable results in practical applications. In the investigation by Södersten and coworkers (2002), the SA channels were so dominated by the subject's voice that even F0 extraction was possible. Instances of phone conversation were erroneously sorted into the background channel, as expected, but since such conversation was not frequent, this posed minor problems. The method also appears to well match a perceptual differentiation of voices recorded with this equipment; a correction of the signals from the Aura program by Aronsson (Unpublished),

performed by hand editing of the SA channel, followed by an F0 and SPL analysis, showed only minute deviations.

In these two studies, the cut-off frequency of the high pass filter was set to 1000 Hz. To improve detection of hypofunctional voices it would be worthwhile to test a HP filter as low as 500 Hz, as suggested by measurements presented above. Another potential improvement is to apply a first order HP filter to the sum channel only rather than a second-order HP filter to both the sum and the difference channels. This would attenuate the SOR at low frequencies, resulting in a 0 dB SOR also for low frequency ambient sounds.

The processing time is obviously considerable. According to the measurements performed on a 1200 MHz AMD Athlon computer, a two-hour long DAT recording would require two hours of transfer to computer file plus a processing time of slightly more than six hours, if all signals were to be extracted. However, there is hardly any need to extract both left and right channels for all signals, and many of the signals are only needed during the development and testing stage or for demonstration purposes. In the study by Södersten et al. (2002), only the SAL, SLL, BAL and BLL channels were extracted, which would reduce the processing time for the two-hour hour tape to 50 minutes on the same computer. Future development of computer technology is likely to further shorten the processing time.

One major difference with the presented recording technique as compared to voice accumulators is that the actual speech of the subject and the persons in his/her surrounding is recorded. This is advantageous in the sense that the processing is performed off-line such that the analysis parameters can be optimised, if needed. Also, results can be examined with respect to reliability. On the other hand, continuous recording of the audio signal may be associated with ethical concerns regarding all recorded speakers. If the main need is quick access to time averages, it would be advantageous to implement the phonation detection strategy described here into future voice accumulators.

In summary, the above method seems useful for an automatic processing of binaural recordings made under realistic conditions. It successfully detects phonation and sorts the recorded sound into separate tracks, one for the speaker's own voice and one for the environmental sound. These tracks can be used for various acoustic measurements, such as phonation time, SPL and F0. Hence, the method should have the potential of significantly improving voice accumulator technology.

Acknowledgements

This work was supported by research grants from the Swedish Council for Work Life Research. I would also like to thank Maria Södersten, Annika Szabo, Carina Aronsson and Johan Liljencrants for valuable discussions and Johan Sundberg for editorial assistance. This work was presented as a poster on the IVth Pan-European Voice Conference (PEVOC IV) and was awarded the "Best Poster Paper in Session One" (of two).

Appendix

Here, the symbols used in the paper are collected and some of the mathematical operations that are performed as well. The signals can be divided into three groups; audio, level and control signals.

	Audio	Level	Control signal
Microphone	M_A	M_L	-
Sum	Σ_A	Σ_L	-
Difference	Δ_A	Δ_L	-
Self	S_A	S_L	S_C
Other	O_A	O_L	O_C
Talk	T_A	T_L	T_C
Background	B_A	B_L	B_C

In addition there is the SOR signal, which usually is expressed as a level. The audio and level signals all exist in a left and right channel version and can be sub-indexed with L or R to indicate left or right channel when necessary.

$\Sigma_A = M_{AL} + M_{AR}$

$\Delta_A = M_{AL} - M_{AR}$

M_L is the level of M_A, Σ_L is the level of Σ_A, Δ_L is the level of Δ_A.
SOR level = $\Sigma_L - \Delta_L$

If S_C is active, $S_L=M_L$ and $S_A=M_A$,
else $S_L=-\infty$ and $S_A=0$

If O_C is active, $O_L=M_L$ and $O_A=M_A$,
else $O_L=-\infty$ and $O_A=0$

If T_C is active, $T_L=M_L$ and $T_A=M_A$,
else $T_L=-\infty$ and $T_A=0$

If B_C is active, $B_L=M_L$ and $B_A=M_A$,
else $B_L=-\infty$ and $B_A=0$

Note: in the original work by Ternström, signals ΣA and ΔA were named “M” and “S”, respectively. The term “Self-to-other ratio” (SOR) was used for the true relation between levels of the own voice and surrounding sounds. Here SOR is used to denote the estimate of the same unless explicitly stated otherwise.

References

Airo E, Olkinuora P, Sala, E (2000). A method to measure speaking time and speech sound pressure level. *Folia Phoniatr Logop* 52: 275-288.

Aronsson C. Evaluation of an automatic phonetogram method using speech material recorded during a working day – a relevant method for vocal loading. (Unpublished, available on request from the author).

Buekers R, Bierens E, Kingma H & Marres EHMA (1995). Vocal load as measured by the voice accumulator. *Folia Phoniatr Logop* 47: 252-261.

Fritzell B (1996). Voice disorders and occupation. *Log Phon Vocol* 21: 7-12.

Kitzing P (1979). Glottographic frequency analysis (in Swedish). *Doctoral dissertation.* Lund University, ENT-clinic, Malmö, Sweden.

Masuda T, Ikeda Y, Manako H & Komiyama S (1993). Analysis of vocal abuse: Fluctuations in phonation time and intensity in 4 groups of speakers. *Acta Otolaryngol* 113: 547-552.

Neils LR & Yairi E (1987). Effects of speaking in noise on vocal fatigue and vocal recovery. *Folia Phoniatr* 39: 104-12.

Ohlsson A-C, Brink O & Löfqvist A (1989). A voice accumulator - validation and application. *J Speech Hear Res* 32: 451-457.

Popolo P, Švec J, Rogge-Miller K & Titze I (2002). Technical considerations in the design of a wearable voice dosimeter. Poster presentation at ASA Cancun, Dec. 2002. *http://192.107.173.4/peter_html/Images/cancun2_112502.pdf*

Rantala L, Lindholm P & Vilkman E (1998). F0 change due to voice loading under laboratory and field conditions. A pilot study. *Log Phon Vocol* 23: 164-168.

Szabo A, Hammarberg B, Håkansson A & Södersten M (2001). A voice accumulator device: Evaluation based on studio and field recordings. *Log Phon Vocol* 26: 102-117.

Szabo A, Hammarberg B, Granqvist S & Södersten M (2003). Methods to study pre-school teachers' voice at work - Simultaneous recordings with a voice accumulator and a DAT recorder. (Accepted for publication in *Log Phon Vocol* 2003).

Södersten M, Granqvist S, Hammarberg B & Szabo A (2002). Vocal behaviour and vocal loading factors for preschool teachers at work studied with binaural DAT recordings. *J Voice* 16/3: 356-371.

Ternström S (1994). Hearing myself with others: Sound levels in choral performance measured with separation of one's own voice from the rest of the choir. *J Voice* 8/4: 293-302.

Ternström S, Södersten M & Bohman M (2002). Cancellation of simulated environmental noise as a tool for measuring vocal performance during noise exposure. *J Voice* 16/2: 195-206.

Titze IR, Švec JG & Popolo PS . Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. *J Speech Lang Hear Res* (Accepted for publication).

van de Weijer J (2002). Terminal rises in infant-directed and adult-directed questions. *Proc Fonetik 2002*, *TMH-QPSR, KTH* 44: 5-8. *http://www.speech.kth.se/qpsr/tmh/2002/02-44-005-008.pdf*

Watanabe H, Shin T, Oda M, Fukaura J & Komiyama S (1987). Measurement of total actual speaking time in a patient with spastic dysphonia. *Folia Phoniatr* 39: 65-70.