Acoustic analyses of postlaryngectomy voice and their perceptual relevance

Corina J van As-Brooks, PhD

Department of Head and Neck Oncology and Surgery, the Netherland Cancer Institute, Amsterdam, The Netherlands

Abstract
Despite the fact that perceptual evaluations are considered the ‘gold standard’ for evaluation of voice quality, there are some drawbacks. They are time-consuming and remain subjective. Hence the use of acoustic measures is interesting. This article describes some common pitfalls of acoustic analyses, the outcomes of acoustic analyses in tracheoesophageal speakers, and their perceptual relevance.

1. Introduction
With the availability of a variety of commercial and free software packages, acoustic analysis has become a tool that is available at any person’s fingertips. However, the acoustic analysis of alaryngeal speech is not straightforward at all and many pitfalls lie ahead. Performing the seemingly easy steps of opening the sound file, running some analyses, and reporting the results may lead to gross errors in the reported results.

The irregularities in periodicity or the total lack of periodicity in many of these voices, can easily cause mistakes in the acoustic analysis. Selection of the type of recording equipment and environment, the speech sample, the software package used, the acoustic parameters analyzed, and checking and correcting the results prior to reporting are all important aspects.

In this paper all these subjects will be addressed.

2. Equipment and Methods

2.1. Recording equipment and environment
When recording alaryngeal voices for acoustic analysis, it is important to obtain recordings of good quality. Therefore, it is always recommended to investigate the current state of the art in regards to recording equipment and the microphone used. Currently, portable digital recorders such as the Edirol RW9 in combination with a good dynamic microphone (for example the Sennheiser MD421 dynamic cardioid microphone) are a good and relatively affordable option.
Recordings maybe made in so-called quite rooms that are treated for optimal acoustic damping or in soundproof booths. The latter obviously provide the highest reduction in noise from the environment.

### 2.2. Speech samples

The type of speech sample that is used may be a sustained vowel (usually the vowel /a/), a (nonsense) syllable or syllable sequence, (nonsense) words or sentences, read-aloud text, retold stories, monologue, dialogue, or conversational speech. The choice of material should fit the goal of the analysis, and often a variety of materials may be used. However, the limitations of acoustic analysis should clearly be kept in mind. For example, F0 and F0 based parameters such as jitter and shimmer can best be measured in the stable part of a sustained /a/ that is produced at a comfortable pitch and loudness. Performing these analyses for example in a vowel produced with the goal to reach maximum phonation time, or in running speech is not advised. Speech samples other than sustained vowels can be useful for studying consonant or vowel characteristics, temporal, or prosodic aspects of alaryngeal speech.

### 2.3. Software programs

Some software packages leave more room for the investigator to change the defaults used for the analysis and adapt them for the alaryngeal speaker, other packages do not allow the investigator to make any changes or only minor changes. For the experienced investigator, the packages that allow the user to change setting are the best option. Some packages also allow the user to adjust for example pitch markers that were placed in error. Commercially available packages are not necessarily better than free packages. The software program Praat is freely available on [www.praat.org](http://www.praat.org) and fulfills all requirements discussed above.

**Figure 1.** Example of pitch marker errors. From: Van As et al. (1998).
3. Results of acoustic analyses in alaryngeal speakers

3.1. Spectrographic analysis

Before carrying out any in-depth acoustic analyses, it is always worthwhile to carry out a spectrographic analysis of the speech sample. A long-term average spectrum, narrow-band spectrogram, and wide-band spectrogram are all useful means of obtaining more insight into the characteristics of the speech signal. The spectrum gives an impression of the harmonicity in the signal, the formant frequencies, and the noise distribution in the signal, the narrow band spectrogram gives an impression of the harmonics and the strength of the harmonics in the signal and can be helpful in determining F0 or checking F0 values reported by the outcomes of the more specific acoustic analyses, the wide-band spectrogram gives an impression of the articulation of consonants and vowels, formant frequencies, and temporal aspects.

The narrow band spectrum can also be used to select the part of the speech sample with the strongest harmonics for further acoustic analysis. In any case, if the spectrographic analysis shows that there are no clear harmonics in the signal, further acoustic analysis should not include any F0-based parameters.

Van As et al. (2006) describe a signal typing system for tracheoesophageal speech that can be used for this purpose.

---

**Figure 1.** Example of a voice sample classified as type I. The oscillogram shows a smooth signal with stable features. The 100 ms selection of the oscillogram shows a clearly periodic pattern. The pitch contour shows a steady fundamental frequency (mean 122 Hz) in the narrow-band spectrogram. Spectral peaks at 1/3 of the pitch frequency are visible. The narrow-band spectrogram of the voice sample even up to 2000 Hz. The long-term average spectrum of 1 s also shows a clear harmonic structure in the lower frequencies and noise in the higher frequency region.

**Figure 2.** Example of a voice sample classified as type II. The oscillogram shows a smooth signal with stable features. The 100 ms selection of the oscillogram shows a nearly periodic pattern, with noise. The pitch contour shows a steady fundamental frequency (mean 70 Hz) in the narrow-band spectrogram. The first harmonic is clearly visible and the second and third harmonics are evident in small peaks of the spectrogram. In the long-term average spectrum of 1 s, only clear three harmonics can be observed and the high-frequency noise is of a higher level than in the type I signal in Figure 1.
The results of acoustic signal typing were linked to the results of an overall judgment of voice quality, and a chi-squared test demonstrated a significant relationship. Type 1 and 2 signal types were more often rated as good voices, whereas type $ signal types were more often rated as poor.

### 3.2. Acoustic parameters

Van As et al. (2006) also described a set of acoustic parameters that can be used in tracheoesophageal speakers. Five of them were fundamental frequency based: median fundamental frequency (F0-MED), standard deviation of fundamental frequency (F0-SD), jitter (JIT), percentage of voiced (%Voiced), and harmonics-to-noise ratio (HNR). Two of them were not: glottal-to-noise- excitation ratio (GNE), and band energy difference (BED). All parameters were calculated using Praat, and settings are described in the article.

As discussed above, one cannot expect to be able to calculate all acoustic parameters for all voice samples of laryngectomized patients. Especially not when the patients were selected randomly and not based on 'good or excellent' voice quality. Van As-Brooks et al. (2006) showed that only 77% of the voice samples could be partially or completely analyzed with respect to the pitch based parameters. In the other samples, visual inspection of the results of the signal typing showed that indeed no clear periodicity was present in the signal. Similarly, Van As et al. reported in 1998, that in about 30% of the voice samples
only very short parts could be analyzed. In a study that investigated the influence of stoma occlusion on voice quality (Van As et al., 1998), 21 patients produced three sustained /a/'s each. Of these samples, one second was selected from the middle of the vowel for further analyses. Complete analysis of all samples was only possible in 62% of the patients. Keeping these limitations in mind, results of acoustic analyses of alaryngeal speech, should always be interpreted with caution. An article reporting these types of results without mentioning any limitations regarding the use of frequency dependent acoustic parameters, is suspicious for overlooking these difficulties and simply reporting the output provided by the software.

A variety of studies is available reporting results of acoustic analysis in alaryngeal speech. Robbins et al. (1984) used measures of frequency, perturbation, and duration in tracheoesophageal, esophageal, and normal speech and found that tracheoesophageal speech is more similar to normal speech than esophageal speech for frequency and duration variables, and that TE speech is louder than esophageal speech. Debruyne et al. (1994) used frequency and perturbation measures, and also frequency-independent measures (spectral slope, harmonic prominence) and found that in TE speech more often a fundamental frequency could be detected and that there was more often a tendency to clearly defined harmonics. Jitter and shimmer values in TE speech were more close to normal speech than those in esophageal speech. Trudeau and Qi (1990) specifically studied male versus female tracheoesophageal speech and concluded that the speech of the women, including pitch, closely resembled that of the men. Van As et al. (1998) reported on TE speech versus normal speech and found that almost all parameters reported on were significantly worse in the TE speakers, and also a much larger variability was found. These results were confirmed in a later study that incorporated the use of signal typing (Van As et al., 2006).

4. Acoustic measures and perceptual relevance

4.1 Perceptual evaluation

Although the acoustic analyses are the main focus of this chapter, they should always be relevant in relation to perceptual evaluations. Obviously, a number generated by an acoustic analysis is only relevant when it expresses a perceptual characteristic of the voice quality. Similar to the acoustic analyses, perceptual evaluations in laryngectomized speakers are not as straightforward as those in normal speakers. The aspects of voice quality that are usually judged for normal pathological laryngeal voices, such a breathiness or hoarseness do often not apply for alaryngeal voice. Conversely, aspects of alaryngeal voices like bubbliness or hypertonicity are not relevant for normal pathological laryngeal voices. Therefore, general perceptual evaluation methods for pathological laryngeal such as GRBAS are often not useful for perceptual evaluation of alaryngeal voice; only the judgment of G (Grade) might be relevant.

In the PhD thesis of Van As (2001) and in Van As et al. (2003) an overview is given that discusses the need for intra- and inter-rater reliability measures; the
factors that influence reliability; the influence of speech materials used; and the various method available. The studies have resulted in a recommendation of the use of naïve raters for gaining insight in daily communication, and trained experts (speech-language pathologists) for gaining insight in clinical performance and for comparison with other clinical measures.

The proposed scales for naïve raters are: ugly-beautiful, deviant-normal, low-high, and deep-shrill and the proposed scales for trained experts are: deviant-normal, ugly-beautiful, breathy-not breathy, hypotonic-not hypotonic, low-high, deep-shrill, slow-quick, and dragging-brisk. Additionally, a judgment of overall voice quality by trained experts, gives a good impression of the voice.

### 4.2 Relationships between perceptual evaluations and acoustic measures

A detailed description of these relationships can be found in Van As (2001) and Van As-Brooks et al. (2006). The acoustic signal types show a significant relationship with overall voice quality, indicating that acoustic signal typing is relevant in relation to overall voice quality.

The correlations between the results of the acoustic measures and the perceptual scales were moderate to strong, showing their relevance with respect to tracheoesophageal voice quality. The complementing scales ugly-beautiful and deviant-normal showed significant correlations with the F0-SD, HNR, %Voiced, and BED. The complementing scales breathy-not breathy and hypotonic-not hypotonic showed significant correlations with HNR, %Voiced, GNE, and BED. The complementing scales low-high and deep-shrill showed a significant correlation with F0-MED. The scales slow-quick and dragging-brisk did not show any relationship with the acoustic measures, which is not surprising as no temporal measures were included in the acoustic analyses.

### 5. Conclusion

If used correctly, acoustic analyses are a good tool for voice quality evaluation of alaryngeal speech. Especially the construction of a spectrum and/or a narrow-band spectrogram gives a very good first impression of the acoustic content of the voice. Based on these, further acoustic analyses may be carried out.

Due to the difficulty in analyzing voices with irregular or absent periodicity, papers reporting on acoustic analysis should always be read critically when it comes to the methodology that was used. Results of papers that presented reliable results show that tracheoesophageal speech is highly variable and remains deviant from normal speech.
6. References


