Comparing Vocal Fold Contact Criteria Derived From Audio and Electroglottographic Signals

[*,†,‡]Laura Enflo, [§,||]Christian T. Herbst, [†]Johan Sundberg, and [*,¶]Anita McAllister, **Linköping, †¶Stockholm, Sweden, ‡Boston, Massachusetts, §Olomouc, Czech Republic, and ||Vienna, Austria*

Summary: Objectives. Collision threshold pressure (CTP), that is, the lowest subglottal pressure facilitating vocal fold contact during phonation, is likely to reflect relevant vocal fold properties. The amplitude of an electroglottographic (EGG) signal or the amplitude of its first derivative (dEGG) has been used as criterion of such contact. Manual measurement of CTP is time consuming, making the development of a simpler, alternative method desirable.

Method. In this investigation, we compare CTP values measured manually to values automatically derived from dEGG and to values derived from a set of alternative parameters, some obtained from audio and some from EGG signals. One of the parameters was the novel EGG wavegram, which visualizes sequences of EGG or dEGG cycles, normalized with respect to period and amplitude. Raters with and without previous acquaintance with EGG analysis marked the disappearance of vocal fold contact in dEGG and in wavegram displays of /pa:/-sequences produced with continuously decreasing vocal loudness by seven singer subjects.

Results. Vocal fold contact was mostly identified accurately in displays of both dEGG amplitude and wavegram. Automatically derived CTP values showed high correlation with those measured manually and with those derived from the ratings of the visual displays. Seven other parameters were tested as criteria of such contact. Mainly, because of noise in the EGG signal, most of them yielded CTP values differing considerably from those derived from the manual and the automatic methods, although the EGG spectrum slope showed a high correlation.

Conclusion. The possibility of measuring CTP automatically seems promising for future investigations.

Key Words: Collision threshold pressure–Vocal fold contact–Wavegram–EGG derivative–EGG spectrum slope–Relative dEGG amplitude.

INTRODUCTION

In song and speech, the voice source is normally created by flow-induced oscillation of the vocal folds. Depending on the amount of airflow and the laryngeal configuration in terms of, for example, the prephonatory glottis, the vocal folds may collide during each cycle of vibration. Vocal fold collision causes an abrupt interruption of transglottal airflow which, in turn, represents the acoustic excitation of the vocal tract; the more abrupt the interruption, the stronger the overtones and the louder the voice produced.

The lowest subglottal pressure needed for vocal fold collision, or the collision threshold pressure (CTP),[1] has been suggested as a complement or an alternative to the phonation threshold pressure (PTP). The latter is defined as the lowest subglottal pressure needed for obtaining and sustaining vocal fold oscillation.[2,3] Accurate measurement of PTP is often difficult due to the very low pressure values[4]; sometimes subjects have difficulties producing extremely soft sounds, and PTP is rather rarely used in clinical practice.[5]

Three previous investigations have explored various factors affecting CTP.[1,6,7] One was concerned with vocal warm-up, one with vocal loading and one with resonance tube phonation with the tube end immersed in water. In these studies, both CTP and PTP were analyzed. It was found that, on average, CTP is about 4 cm H_2O higher and has a lower coefficient of variation than PTP.[1] Furthermore, CTP and PTP tended to be lower after vocal warm-up.[1] Both after vocal loading and resonance tube phonation with the tube end in water, CTP was found to rise significantly.[6,7] In the vocal loading experiment also, the PTP increase was significant.

Measurement of CTP requires information about both vocal fold contact and subglottal pressure. Knowledge of the former can be obtained from electroglottographic signals. Electroglottography, henceforth EGG, is a commonly used, noninvasive tool for clinical and research purposes. This signal has been found to be related to vocal fold contact area, although it does not offer an absolute measure of it; thus, an increase of EGG merely reflects an increase of glottal contact area.[8]

Phonation produced with vocal fold contact generally generates larger EGG amplitudes than phonation without such contact.[8] This amplitude was used for CTP determination (Figure 1).[1,7] Another indicator of vocal fold contact is a marked knee in the EGG waveform, so the derivative of the EGG signal, henceforth dEGG, is useful as a criterion of vocal fold contact.[9–11] Disappearance of spikes in the dEGG signal would signify loss of vocal fold contact. The dEGG parameter was used in a previous CTP investigation.[6] An example of dEGG is shown in Figure 1.

The recently invented wavegram offers a visual representation of waveform properties and is potentially useful as a tool for assessing the presence of vocal fold contact.[12] In a wavegram, the EGG or dEGG waveform is decomposed into individual cycles, each normalized in both duration and

Accepted for publication May 20, 2015.

From the *Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden; †Department of Speech, Music and Hearing, Royal Institute of Technology (KTH), Stockholm, Sweden; ‡Department of Speech, Language and Hearing Sciences, Boston University, Boston, Massachusetts; §Voice Research Laboratory, Department of Biophysics, Faculty of Science, Palacký University Olomouc, Olomouc, Czech Republic; ||Laboratory of Bio-Acoustics, Department of Cognitive Biology, University of Vienna, Vienna, Austria; and the ¶Department of CLINTEC, Karolinska Institute, Stockholm, Sweden.

Address correspondence and reprint requests to Laura Enflo, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden. E-mail: lenflo@kth.se

Journal of Voice, Vol. 30, No. 4, pp. 381-388
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2015.05.015

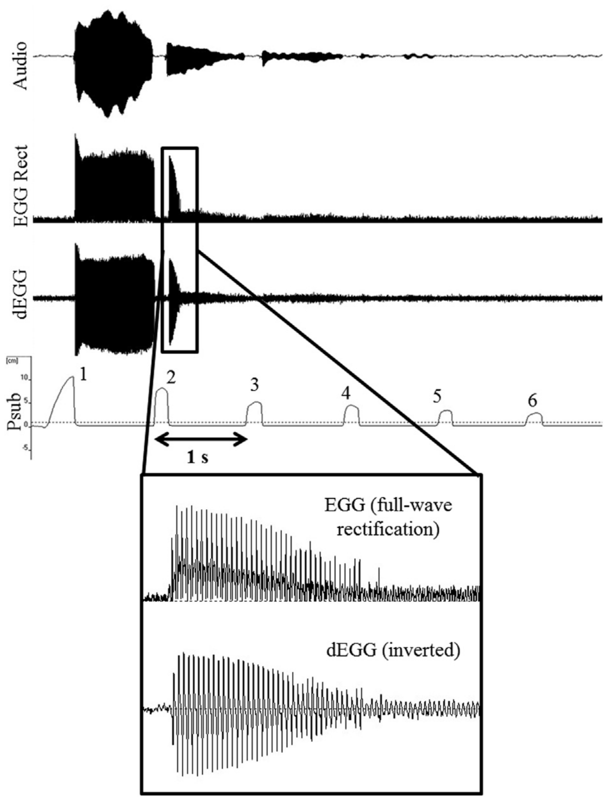

FIGURE 1. Signals used for detection of loss of vocal fold contact in a /pa:/-sequence sung by a female singer at E4 (329.6 Hz). The *top curve* shows the audio signal, the *second and third curves* from the top shows a full-wave rectification of the EGG and the inverted dEGG, respectively. The *fourth curve* from the top shows the orally estimated P_{sub}. The *bottom panel* is a close-up of the EGG and dEGG for the vowel segment in the second /pa:/ in the sequence. The mean of pressure peaks 2 and 3 seen in the *fourth curve* from the top is an estimation of CTP.

amplitude. The amplitude of the waveform within these cycles is then color-coded, and the resulting strips of pixels are consecutively displayed from bottom to top along the x-axis, representing overall time. The color intensity corresponds to the instantaneous EGG or dEGG signal amplitude within each cycle. Disappearance of vocal fold contact is shown as a change of the wavegram pattern.

Subglottal pressure, henceforth P_{sub}, is difficult to measure directly because that requires percutaneous insertion of a pressure transducer into the trachea.[13] A commonly used alternative noninvasive method, used in all our previous CTP studies, is to measure the oral pressure during /p/ occlusion.[14,15] Here, P_{sub} is recorded by means of a pressure transducer attached to a thin plastic tube which the subject holds in the corner of the mouth.

Measurement of CTP requires simultaneous recordings of audio, EGG, and P_{sub} while the subject phonates a series of /pa:/-sequences, starting at medium vocal loudness and continuing until phonation ceases. CTP is approximated as the mean value of the pressure peak preceding and following an abrupt decrease of EGG amplitude or disappearance of dEGG spikes.

This study examines a set of alternative criteria of vocal fold contact derived from EGG data and from the audio signal. They are compared against results of a visual inspection of the dEGG signal, previously found to be reliable.[6] On the basis of these comparisons, a semi-automatic method for identifying vocal fold contact is developed. The corresponding pressure values are then identified and expressed as CTP values.

METHOD

Recordings made in earlier investigations were used in this study. The recording conditions have been described in detail elsewhere.[1,6,7] P_{sub} was recorded by means of a pressure transducer (Gaeltec Ltd, 7b, Isle of Skye, UK) that the participants held in the corner of their mouth. A pressure calibration signal, determined by means of a manometer, was recorded for each singer. The subjects, four female and three male singers, all experienced soloists or choristers, produced series of /pa:/-sequences in *legato* (evenly tied together) and *decrescendo* (gradually decreasing vocal loudness) on various fundamental frequencies ($F0$), chosen according to the personal vocal range, starting at medium loudness and continuing until phonation ceased. The rationale for using data from singers only was so as to minimize variations in $F0$ due to changes in vocal loudness for each /pa:/-sequence, which, if such problems would occur, might make correct CTP estimations more difficult to obtain. At least three /pa:/-sequences were recorded at each $F0$. All recordings were made with the singer in sitting position in a sound-treated booth, having a background noise level 33 dB(A), 40 dB(C).

Each /p/ and each /a/ in the /pa:/-sequences was segmented manually. An automatic method was developed for measuring the pressure during the /p/ segments. This method identified and returned the peak P_{sub} amplitude in each /p/ utterance. Mostly, the pressure peaks were flat, thus yielding reliable data.[15,16] However, in approximately 7% of the cases, the peaks were pointed, suggesting a changing pressure during the occlusion or some other complication. This often occurred at high $F0$. In such cases, no reliable pressure values could be obtained. In the evaluation of each parameter, only reliable P_{sub} values were used. As a complement and control of these automatically measured P_{sub} values, also manual measurements were carried out.

In a decrescendo sequence, P_{sub} is likely to decrease continuously. The mean duration of the /pa:/-sequences was typically in the vicinity of 5 seconds. In this time interval, the subjects mostly produced seven to ten syllables, each having a duration of 700–500 ms. During the crescendo, the subjects mostly reduced their P_{sub} by about 5% per /pa:/ syllable.

EGG signals sometimes contain low-frequency noise of substantial amplitude, typically due to larynx movement. Hence, using the Extract tool in the Soundswell Signal Workstation (Hitech Medical, Saven Hitech AB, Stockholm, Sweden), all recorded EGG signals were first high-pass filtered at 90 Hz, applying zero phase filter setting. Then, the resulting signal was differentiated, again using the same Extract tool.

Using the dEGG signal as criterion of vocal fold contact, the CTP values were estimated as the average of the two pressure