Flow Glottogram Characteristics and Perceived Degree of Phonatory Pressedness

*Moa Millgård, *Tobias Fors, and †,‡Johan Sundberg, *Gothenburg, †‡Stockholm, Sweden

Summary: Objectives. Phonatory pressedness is a clinically relevant aspect of voice, which generally is analyzed by auditory perception. The present investigation aimed at identifying voice source and formant characteristics related to experts' ratings of phonatory pressedness.

Study Design. Experimental study of the relations between visual analog scale ratings of phonatory pressedness and voice source parameters in healthy voices.

Methods. Audio, electroglottogram, and subglottal pressure, estimated from oral pressure during /p/ occlusion, were recorded from five female and six male subjects, each of whom deliberately varied phonation type between neutral, flow, and pressed in the syllable /pae/, produced at three loudness levels and three pitches. Speech-language pathologists rated, along a visual analog scale, the degree of perceived phonatory pressedness in these samples.

Results. The samples were analyzed by means of inverse filtering with regard to closed quotient, dominance of the voice source fundamental, normalized amplitude quotient, peak-to-peak flow amplitude, as well as formant frequencies and the alpha ratio of spectrum energy above and below 1000 Hz. The results were compared with the rating data, which showed that the ratings were closely related to voice source parameters.

Conclusions. Approximately, 70% of the variance of the ratings could be explained by the voice source parameters. A multiple linear regression analysis suggested that perceived phonatory pressedness is related most closely to subglottal pressure, closed quotient, and the two lowest formants.

Key Words: Voice source–Phonation type–Subglottal pressure–Flow glottogram–Perceived pressedness.

INTRODUCTION

Hyperadduction or phonatory pressedness is known as a risk factor for vocal health, often causing voice disorders.[1] It affects the perceived voice quality and is therefore included as an important parameter in perceptual voice evaluation systems, for example, SVEA (Stockholm Voice Evaluation Approach), CAPE-V(The Consensus Auditory-Perceptual Evaluation of Voice), and GRBAS (Grade, Roughness, Breathiness, Asthenia, and Strain).[2–4] The differences between these evaluation systems of course represent a serious limitation. Furthermore, perceptual evaluations are time-consuming and somewhat subjective.[5] Therefore, methods based on physical data would be desirable at least as a complement and, in the future, as a replacement of the currently used evaluation methods.

Hyperfunction refers to a characteristic of the voice source, that is, of the pulsating glottal airflow. The main physiological control parameters of the voice source are subglottal pressure, vocal fold length and tension and glottal adduction, controlling vocal loudness, pitch, and type of phonation.[6] The effects on the voice source due to variation of these parameters are complex. For example, firm glottal adduction must be combined with an elevated subglottal pressure, which in turn tends to raise the fundamental frequency.[7] Furthermore, an increase of the sound level of a radiated vowel, that is, an increase of the subglottal pressure is typically associated with several effects: (1) a change in the spectral balance between high and low partials, (2) an increase of the relative duration of the closed phase, (3) a decrease of the level difference between the first and the second voice source partial, and (4) an increase of the peak-to-peak flow amplitude.[8]

It is a well-established fact that the voice source can be estimated by inverse filtering the flow signal, obtained from a flow mask.[9] As flow equals the derivative of pressure, the flow signal can be derived from the audio signal. Inverse filtering analysis implies that the input signal is filtered by the inverted frequency response curve of the vocal tract. Thus, the frequencies and bandwidths of the vocal tract resonances, that is, the formant frequencies and bandwidths, are determined, and the filter is adjusted to these values. The result is a flow glottogram showing glottal airflow versus time, being a graphical representation of the sound injected into the vocal tract. A flow glottogram is physiologically realistic in the sense that it reflects the duration of the closed phase, the airflow during the closed phase (if any), and the amount and changes of the glottal airflow during the vibratory cycle.[10]

The aim of the present study was to identify acoustic and/or physiological correlates of perceived pressedness. Thus, the question we asked was which physical parameters, or combination of parameters, predict ratings of perceived phonatory pressedness?

METHODS

Recordings

Recordings were made of 11 healthy speakers (five women and six men, age range 28–74 years, mean 39 years) who all volunteered to participate. They were all able to vary vocal loudness

Accepted for publication March 23, 2015.

The work was presented at the Annual Symposium Care of the Professional Voice, Philadelphia, June 2012 and at the fifth International Congress of World Voice Consortium, Luxor, Egypt, October 2012.

From the *Division of Speech and Language Pathology, Institute of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden; †Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH, Stockholm, Sweden; and the ‡University College of Music Education, Stockholm, Sweden.

Address correspondence and reprint requests to Johan Sundberg, Department of Speech, Music and Hearing, School of Computer Science and Communication, KTH, Stockholm, Sweden. E-mail: pjohan@speech.kth.se

Journal of Voice, Vol. 30, No. 3, pp. 287-292

0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2015.03.014

independent of pitch. None reported voice problems at the time of the recording.

Subjects were instructed to repeat the syllable /pæ:/ in legato at least four times at medium, low, and high pitch in neutral, soft, and loud voice. The syllables were about 700 ms long. The three pitches corresponded to the subjects' normal speaking pitch, five or seven semitones lower, and five or seven semitones above speaking pitch. The latter two pitches were presented to the subjects by means of the Madde vowel synthesizer software written by Svante Granqvist, KTH (available at www.tolvan.com, last inspected 150315). The subjects produced this exercise in neutral and pressed type of phonation. If participants were unfamiliar with the latter condition, it was demonstrated to them by of one of the authors. Five subjects were acquainted with how to produce also flow phonation, defined as the phonation which has the lowest degree of glottal adduction that still results in vocal fold contact.[11] These subjects repeated the exercise also in this type of phonation. Thus, six participants phonated under 18 different conditions (two phonation types × three pitches × 3° of vocal loudness) and five participants under 27 conditions (three phonation types × three pitches × three degrees of vocal loudness), such that a total of 243 voice samples were collected. Before the recording, subjects were informed about the purpose of the study and the recording procedure and gave their written consent to participate.

The recordings were made in a sound-treated room (5 × 4 × 2.5 m). Because flow mask recordings of airflow have a limited frequency range, the audio signal was preferred for voice source analysis. Using the Soundswell signal workstation software (Hitech Development, Solna, Sweden), audio, electroglottogram (EGG), and pressure were simultaneously recorded on separate channels in the SMP format. Audio was captured by means of a head-worn omnidirectional electret DPA 4066-C microphone (DPA Microphones, Allerod, Denmark) located at a measured distance from the mouth. The microphone signal was amplified by a Symetrix SX202 amplifier (Symetrix Inc., W. Lynnwood, WA). The signal was calibrated by means of a 1000-Hz sine wave, the sound pressure level (SPL) of which was measured by a sound level meter at the recording microphone; the SPL value was announced in the recording. EGG was obtained from a Glottal Enterprises (Syracuse, NY) MC2-1 dual channel device. Oral pressure was measured by means of a thin plastic tube, attached to a pressure transducer included in the Glottal Enterprises, XMSIF-2 equipment. The pressure signal was calibrated by recording pressure values determined by means of a manometer. These pressure values were announced in the recording. The subject held the tube ending in the corner of the mouth. During the recordings, the EGG and pressure signals were monitored on an oscilloscope.

Analyses

Perception. To obtain quantitative data about perceived degree of pressedness, a listening test was run. From each subject, three neutral and three pressed samples were randomly selected at each pitch, and, in addition, three flow phonation examples were randomly selected from the five subjects who had produced also such examples. This yielded a total of 243 samples (three pitches × three loudnesses × two phonation types × six subjects; three pitches × three loudnesses × three phonation types × five subjects). To limit the number of stimuli in the test, only 79 of these syllables were selected.

Each syllable was presented three times, separated by a pause of 150 ms. Using the Internet freeware *SurveyGizmo* (Boulder, CO, www.surveygizmo.com, last inspected 141028), the stimuli were organized into files, each including the 79 stimuli plus 20 replicated stimuli, randomly selected. These stimuli were presented in a randomized sequence unique to each listener. The test was distributed over the Internet.

The listeners' task was to rate the perceived degree of pressedness along a visual analog scale (VAS) that appeared on the computer screen. The left and right extremes of the "hyperfunction/press" VAS were marked "none" and "extreme." At the beginning of the test, five practice stimuli were presented, the ratings of which were discarded. Because the test was distributed over the Internet, the listeners ran the session on their own and could take a break whenever they liked. Also, they could listen to a given stimulus any number of times, but they could not return to an earlier stimulus. The listeners were instructed to use high-quality earphones and were asked to specify the type of earphones that they used. The test was run by 16 speech-language pathologists, all with professional experience of voice, and the results were recorded by the *SurveyGizmo* software.

Acoustics. The calibrated audio signal was inverse filtered using the custom made *Decap* software (Svante Granqvist, KTH, www.tolvan.com, last inspected 150315). This program displays waveform and spectrum in separate windows (Figure 1). It converts the input audio signal to flow by integration and applies the classical equations for calculating the transfer function corresponding to the manually adjusted

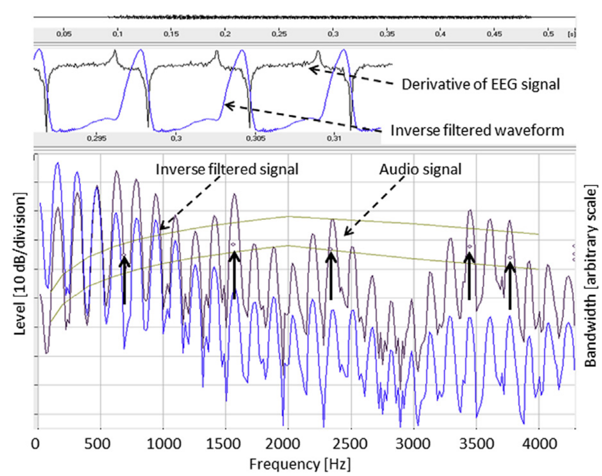

FIGURE 1. Display of *Decap* software analysis of a pressed sample produced by male subject 3. From *top* to *bottom*, the windows represent the audio; the waveforms of the derivative of the EGG signal and the inverse filtered audio signal; and the spectrum of the audio signal and the inverse filtered audio signal. The *arrows* show the frequencies and, along an arbitrary scale, the bandwidths of the inverse filters; the *curves* represent typically occurring bandwidth values.