Publication VII

Paavo Alku, Matti Airas, Eva Björkner, and Johan Sundberg, "An amplitude quotient based method to analyze changes in the shape of the glottal pulse in the regulation of vocal intensity." *Journal of the Acoustical Society of America*, 120(2), pp. 1052–1062, August 2006.

An amplitude quotient based method to analyze changes in the shape of the glottal pulse in the regulation of vocal intensity

Paavo Alku[a)] and Matti Airas
Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, P.O. Box 3000, Fin-02015 TKK, Finland

Eva Björkner and Johan Sundberg
Department of Speech, Music and Hearing, Royal Institute of Technology, Lindstedtsvägen 24, SE-10044, Stockholm, Sweden

(Received 16 February 2006; revised 10 May 2006; accepted 17 May 2006)

This study presents an approach to visualizing intensity regulation in speech. The method expresses a voice sample in a two-dimensional space using amplitude-domain values extracted from the glottal flow estimated by inverse filtering. The two-dimensional presentation is obtained by expressing a time-domain measure of the glottal pulse, the amplitude quotient (AQ), as a function of the negative peak amplitude of the flow derivative (d_{peak}). The regulation of vocal intensity was analyzed with the proposed method from voices varying from extremely soft to very loud with a SPL range of approximately 55 dB. When vocal intensity was increased, the speech samples first showed a rapidly decreasing trend as expressed on the proposed AQ-d_{peak} graph. When intensity was further raised, the location of the samples converged toward a horizontal line, the asymptote of a hypothetical hyperbola. This behavior of the AQ-d_{peak} graph indicates that the intensity regulation strategy changes from laryngeal to respiratory mechanisms and the method chosen makes it possible to quantify how control mechanisms underlying the regulation of vocal intensity change gradually between the two means. The proposed presentation constitutes an easy-to-implement method to visualize the function of voice production in intensity regulation because the only information needed is the glottal flow wave form estimated by inverse filtering the acoustic speech pressure signal. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2211589]

PACS number(s): 43.70.Jt, 43.70.Gr [BHS] Pages: 1052–1062

I. INTRODUCTION

The regulation of vocal intensity is an inherent feature in speech communication. Our everyday life is full of events in which we want to vary the volume of speech: in order to emphasize something, in order to be heard over a longer distance, in noisy environments, or, for example, in order to convey emotional content (such as anger) in the speech message. Intensity of sound is defined in physics as a measure of power per unit of area (e.g., Young and Freedman, 1996). In speech science, the term *vocal intensity* is commonly used to refer to the acoustic energy of speech, and it is typically quantified by using the sound pressure level (SPL) (Titze, 1994). Healthy speakers are able to produce voiced sounds over a wide range of intensity levels. In a study by Coleman *et al.* (1977), for example, voices of female and male subjects were analyzed by having each subject produce the maximum and minimum SPL of which he or she was capable. The values obtained, measured at a distance of 6 in. from the speaker, showed that the speech samples produced in soft and (very) loud phonation varied between 48 and 126 dB.

Changing sound intensity in audio equipment corresponds simply to amplitude scaling, i.e., multiplication with a constant that is either larger than unity (amplification) or smaller than unity (attenuation). In contrast to this, intensity regulation of natural speech uses various mechanisms realized by the human voice production apparatus. Consequently, varying speech intensity is reflected in changes of many acoustical features of the speech wave form, such as the fundamental frequency ($F0$), formant values, and the spectral tilt of the produced speech sound, rather than in pure amplification/attenuation of the signal wave form alone. According to Titze (1994), there are three basically distinct mechanisms to control the intensity of the human voice. They correspond to adjustments of the vocal apparatus below the larynx, within the larynx, and above the larynx. Below the larynx, intensity can be regulated by controlling the aerodynamic output of the lungs to the vocal system. The key variable in this process is subglottal pressure[1] (Ladefoged and McKinney, 1963; Bouhuys *et al.*, 1968). Within the larynx, there are methods to regulate intensity by modifying the vibration of the vocal folds and hence changing the amount of aerodynamic power converted into acoustic power. These methods correspond, for example, to an increase in the flow amplitude or to a decrease in the length of the glottal closing phase and they are typically reflected by a changed value of subglottal pressure. Above the larynx, vocal intensity can be modified by adjusting the resonances of the vocal cavity, especially the first formant, to coincide with the harmonics

[a)]Electronic mail: paavo.alku@tkk.fi

 0001-4966/2006/120(2)/1052/11/$22.50

of the glottal source. This method to control vocal intensity is called formant tuning, but it is not typically used in intensity regulation of conversational speech (Titze, 1994).

In the late 1950s and early 1960s, a number of investigations were carried out on the intensity regulation of speech (Van den Berg, 1956; Ladefoged and McKinney, 1963; Isshiki, 1964, 1965). These studies addressed mainly the impacts of subglottal pressure and air flow rate on vocal intensity. The key role of the subglottal pressure on vocal intensity was demonstrated, for example, in a study by Ladefoged and McKinney (1963), who reported that the peak sound pressure level (in dB) rose linearly as a function of subglottal pressure (presented on the logarithmic scale) for the middle part of the subglottal pressure range. On the other hand, Isshiki (1964, 1965) studied intensity regulation as controlled by two factors, the contraction of the laryngeal adductor muscles and the expiratory force. He reported that in low frequency phonation the flow rate remained almost unchanged, or even slightly decreased, with the increase in voice intensity reflecting the use of laryngeal control. In contrast to this phenomenon, the flow rate of high frequency phonation showed a large increase when voice intensity was raised indicating the use of expiratory muscle control. These results were supported in a laryngeal electromyography study by Hirano *et al.* (1970). Subsequently, Isshiki's approach to the study of vocal intensity based on respiratory and laryngeal phenomena has been used in various investigations. Stathopoulos and Sapienza (1993a, 1993b) for example, compared respiratory and laryngeal function in intensity variation between males and females and between adults and children. In addition, the two factors suggested by Isshiki have been used in studies comparing intensity regulation of normal and disordered voices (e.g., Makiyama *et al.*, 2005) as well as in studies addressing sustained and transient increases of intensity (Finnegan *et al.*, 2000). Even though the roles of the laryngeal adductor muscles and the expiratory force have been studied widely since they were originally proposed by Isshiki, it is worth noting that there is still a lack of understanding on how these two major factors interact when vocal intensity is regulated.

The early studies on intensity regulation of speech did not, however, address the relationship between the excitation of (voiced) speech created by the vibrating vocal folds, the glottal volume velocity wave form, and vocal intensity. Later, when the use of inverse filtering became widely used, numerous experiments were conducted to study the behavior of the glottal volume velocity wave form in intensity regulation of speech. These studies show, in terms of the amplitude-based parameters of the glottal flow (Fig. 1), that raising the vocal intensity is accomplished typically by an increase of the ac flow and also (but perhaps not so clearly) by a decrease of the minimum flow (Holmberg *et al.*, 1988; Hertegård and Gauffin, 1991; Titze, 1992b; Sulter and Wit, 1996). Furthermore, increasing vocal intensity corresponds to changing the shape of the glottal pulse so that the length of the open phase decreases (Dromey *et al.*, 1992; Sapienza *et al.*, 1998; Hodge *et al.*, 2001) and, in particular, the length of the glottal closing phase decreases (Holmberg *et al.*, 1988; Sulter and Wit, 1996). The increase of the ac flow and the

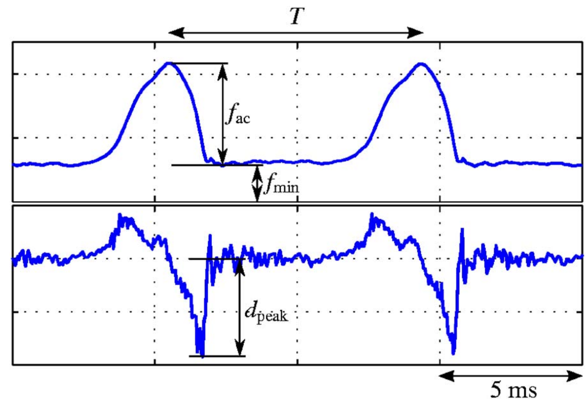

FIG. 1. (Color online) Glottal flow pulse (upper graph) and its first derivative (lower graph) computed from natural speech using inverse filtering. The amplitude-domain quantities shown as: f_{ac}=peak-to-peak flow (also called the ac flow), f_{min}=minimum of the flow (also called the dc offset), d_{peak}=negative peak amplitude of the flow derivative (also called the maximum airflow declination rate). Length of the fundamental period is denoted by T.

decrease of the duration of the glottal closing phase both amplify the negative peak amplitude of the flow derivative (Sundberg *et al.*, 1993). This parameter, depicted by d_{peak} in Fig. 1, is also called the maximum flow declination rate (Holmberg *et al.*, 1988; Stathopoulos and Sapienza, 1993a, 1993b). It has been shown in various studies (e.g., Holmberg *et al.*, 1988; Gauffin and Sundberg, 1989; Fant, 1993; Stathopoulos and Sapienza, 1993b; Sapienza and Stathopoulos, 1994; Sulter and Wit, 1996; Alku *et al.*, 1999) that d_{peak} is closely related to vocal intensity. The important role of d_{peak} characterizing the behavior of the voice source in intensity regulation is demonstrated by the fact that the mapping between SPL (in dB) and d_{peak} (on the logarithmic scale) can be approximated very closely with an increasing linear function (Gauffin and Sundberg, 1989; Alku *et al.*, 1999).

Titze and Sundberg studied the regulation of vocal intensity in speakers and singers by presenting analytic and empirical models to predict sound pressure levels from glottal wave form parameters (Titze, 1992a; Titze and Sundberg, 1992). In their studies, the role of minimum lung pressure to initiate phonation, the phonation threshold pressure, was essential. The authors addressed, among other things, the question of how subglottal pressure affects the vocal intensity. It was shown that the doubling of excess pressure over the threshold caused an 8–9 dB increase in SPL. The first 6 dB of this increase was explained by an increase in the glottal flow amplitude, whereas the remaining 2–3 dB came from changing the shape of the flow pulse to a less symmetric form. Similar results were also reported by Fant (1982), whose theoretical calculations yielded the SPL increase of 9.5 dB as a result of doubling subglottal pressure. According to Fant, 2 dB of this rise in vocal intensity is caused by changes in the glottal flow pulse form during the closing phase.

The relationship between vocal intensity and $F0$, or its perceptual counterpart, loudness, has also been studied widely (e.g., Gramming *et al.* 1988; Strik and Boves, 1992; Titze and Sundberg, 1992; Alku *et al.*, 2002). Attention has been focused especially on the voice range profile (VRP), also called the phonetogram, which is a display of vocal intensity range versus $F0$ (Damste, 1970; Gramming and Sundberg, 1988; Åkerlund *et al.*, 1992; Titze, 1992b; Coleman, 1993). In one of their VRP studies, Gramming *et al.* (1988) reported that the mean pitch increased by about a half-semitone when intensity was increased by one decibel. According to their findings, the increased value of $F0$ can be expected as a passive result of raising the subglottal pressure in order to produce louder sounds.

The present study addresses the question of how speakers change the characteristics of the glottal pulse when vocal intensity is increased. Given the essential role of intensity regulation in voice communication, this issue is undoubtedly of great importance in speech science. Previous studies, however, have paid little attention to methodologies to visualize how the glottal pulse characteristics change when intensity is regulated. Therefore, a new method is presented which makes it possible to analyze how a rise in vocal intensity is contributed to by the following two glottal source related mechanisms both of which raise the negative peak amplitude of the flow derivative: (a) changing of the glottal pulse to a more asymmetric form versus (b) increasing of the glottal flow amplitude. Rather than using the ordinary approach (e.g., Holmberg *et al.*, 1988; Stathopoulos and Sapienza, 1993a, 1993b; Södersten *et al.*, 1995; Sulter and Wit, 1996), in which speech samples are divided into few intensity categories (typically three: soft, normal, loud), this study analyzes voices, the intensity of which rose from low to high in small steps of approximately 5 dB. This approach, previously used also by Sundberg *et al.* (2005), was chosen because it allows a more detailed insight into the intensity regulation mechanisms in speech. Furthermore, many of the previous studies on intensity regulation of speech have used a rather narrow SPL range between soft and loud phonations. For example, the difference between loud and soft male voices was, on average, 11 dB in the study by Holmberg *et al.* (1988), 15 dB in the study by Sapienza *et al.* (1998), and 18 dB in the study by Hodge *et al.* (2001). In contrast to these studies, the present survey aimed to gather information on vocal intensity regulation mechanisms by using phonations that had SPL dynamics much closer to the extreme limits achievable by normal speakers.

II. MATERIAL AND METHODS

A. Speech material

Speech data were-collected from ten adult Finnish speakers (five females and five males) with no history of speech, voice, or hearing disorders. The ages of the female speakers varied between 42 and 54 and those of the males between 33 and 45 years. The acoustic speech pressure wave form was recorded using a condenser microphone (Brüel&Kjaer 4176) placed at a distance of 40 cm from the lips of the speaker. The distance was carefully controlled before each phonation, an important precaution in order to be able to compare the amplitude features of the estimated glottal flows between different phonations.(see Sec. II D). Intraoral pressure was simultaneously recorded using a pressure transducer (Frøkjær-Jensen Electronics, Manophone, MF710) connected to the subject's mouth with a plastic catheter. For the computation of SPL values, we also recorded a calibration signal generated by a Brüel&Kjaer 4231 calibrator. During the recording, the subject phonated with the catheter placed between the lips in the corner of the mouth. The recordings were made in an anechoic chamber and all of the subjects sat while producing the sounds.

The speaking task was to produce a series of the word /pa:p:a/ with increasing vocal intensity. First, each subject was asked to repeat the production of the word /pa:p:a/ several times as softly as possible without whispering. The output level of the speech signal was measured by means of a sound level meter (Brüel&Kjaer 2225), whose LED display was visible to both the subject and the experimenter. When the subject had learned to repeat the production of the test word as softly as possible with a consistent SPL value, the recording of the first sample was conducted. The subject was then asked to produce the word /pa:p:a/ by increasing the sound pressure level by 5 dB. In order to do this, the subject again watched the LED display of the sound level meter and gave multiple productions until the desired SPL increase from the preceding representation was achieved. When the desired SPL value was reached, five repetitions of the /pa:p:a/ word were recorded. The same procedure was then repeated several times with the subject increasing the value of SPL by 5 dB until she or he reached the loudest sample required with a SPL value of 105 dB. (Some subjects also voluntarily produced an even louder sound with an SPL value of 110 dB.) All the subjects found this procedure straightforward and they achieved the desired SPL value after a few productions of the word /pa:p:a/. Importantly, the subjects were given no other restrictions regarding their voice production, permitting the speakers to choose pitch and phonation type freely during the recording.

Both the acoustic speech pressure wave form and the intraoral pressure signal were directly transmitted to a computer via an audio card (Turtle Beach Tahiti, Turtle Beach Systems) with a frequency response from dc up to 20 kHz. Each recorded speech sample comprised five productions of the /pa:p:a/ repetitions, from which the middle one was always selected and stored for further analyses. The wave forms were digitized using a sampling frequency of 22.050 kHz and a resolution of 16 bits. The bandwidth of the signals was 11 kHz. In the computer, the acoustic speech signals were first high-pass filtered with a linear phase FIR filter, with a cut-off frequency of 60 Hz, in order to remove any possible low-frequency air pressure variations picked up during the recordings.

B. Inverse filtering

In order to estimate the glottal volume velocity wave forms, we used an inverse filtering technique, which estimates the glottal excitation directly from the acoustic speech pressure wave form as recorded in a free field, i.e., no flow mask was used. The method used was a slightly modified version of the technique described by Alku (1992). The modification concerned the modeling of the vocal tract transfer function, which was based on an all-pole modeling technique, called Discrete All-Pole Modeling (DAP) (El-Jaroudi and Makhoul, 1991), instead of the conventional Linear Predictive Coding (LPC) used by Alku (1992). The difference between DAP and LPC is that the former is based on the Itakura-Saito distortion criterion in determining an optimal all-pole filter, whereas the latter uses the least squares error criterion. Consequently, the formants of the vocal tract, particularly for high-pitched voices can be more accurately estimated by DAP than by LPC (El-Jaroudi and Makhoul, 1991). An accurate modeling of the vocal tract transfer function is crucial from the point of view of glottal inverse filtering. As reported by Alku and Vilkman (1994), the application of DAP instead of LPC in modeling the vocal tract transfer function decreases the amount of formant ripple in the estimated glottal flows.

The inverse filtering method used in the present study estimates the vocal tract transfer function applying the above-mentioned DAP technique by first canceling the average effect of the glottal source from the speech spectrum using low-order all-pole filtering. In order to be able to compare the different phonations in terms of the value of d_{peak}, a normalization method described in detail in Alku *et al.* (1998) was used. In this method, the dc gain of the vocal tract model used in inverse filtering is adjusted to unity. This makes possible an analysis of the amplitude features of the glottal flow on an amplitude scale, which is arbitrary (i.e., it does not yield the absolute flow values), but remains the same for all the voices analyzed. The length of the analysis window was 50 ms. For some of the low-pitched male voices, the length of the time window was increased to 70 ms in order to cover at least four glottal cycles. In order to estimate the glottal flow over sustained phonation, the analysis window of inverse filtering was set at the middle of the long vowel in the/pa:p:a/word.

C. Computation of sound pressure level and subglottal pressure

The SPL values were determined on the dB scale using the energy computation[2] and the SPL value of the calibration tone (94 dB) as follows:

$$L_p = 94\ \text{dB} + 10\log_{10}\frac{E_s}{E_c}, \tag{1}$$

where L_p is speech *SPL*, E_s is the energy of the speech signal, and E_c is the energy of the equally long calibration signal. It is worth noting that Eq. (1) here yields the SPL value of a speech signal with linear weighting at a distance of 40 cm from the speaker's lips. In order to be able to compare the results of inverse filtering to SPL values, the analysis window of the SPL computation was adjusted to start from the same position as in inverse filtering. However, the window length in SPL computation was longer (100 ms) than that of inverse filtering.

Estimates of subglottal pressure (P_s) were computed from the oral-pressure measurements, a practice widely used in voice production studies. In this approach, the subglottal pressure of a vowel is estimated from the oral-pressure measurements during the stop consonants surrounding the vowel.

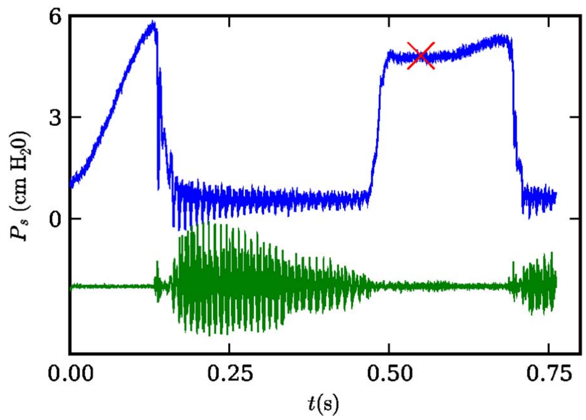

FIG. 2. (Color online) Example of the oral-pressure (upper graph) and the speech pressure wave form (lower graph) phonated by a male subject. Two magnified sections of the oral-pressure signal correspond to the /p/-occlusions of the /pa:p:a/ word. The second of these, at the instant indicated by a cross, was used for the computation of the subglottal pressure value.

If several steady-state pressure values are available, it is possible to use interpolation of consecutive oral-pressure values (e.g., Löfqvist *et al.*, 1982; Holmberg *et al.*, 1988; Hertegård *et al.*, 1995). In the present study, however, there were only two /p/-occlusions, the prevocalic and postvocalic /p/-occlusion of the /pa:p:a/ word, available in the digital signals stored in the computer. From these two measurements, the subglottal pressure was estimated by taking into account only the postvocalic /p/-occlusion because it was found to be more stable. The value of P_s was estimated by searching for the "knee" of the oral-pressure wave form during the postvocalic /p/-occlusion (see Fig. 2). The "knee" was determined subjectively by the experimenter by searching for the point in the rising phase of the oral-pressure signal at which the signal reached a plateau level (a similar procedure has been used, for example, in Hertegård *et al.*, 1995).

D. Amplitude quotient as a function of the negative peak amplitude of the glottal flow derivative

In the quantification of the time-based features of the glottal flow it has become accepted that these are computed by measuring the *time intervals* between certain events such as glottal opening and closure as well as the instant of the maximal flow (Holmberg *et al.*, 1988; Cummings and Clements, 1995; Sulter and Wit, 1996; Vilkman *et al.*, 1997). However, it is also possible to measure time-domain features of the glottal closing phase using the *amplitude-domain* values extracted from the glottal flow and its first derivative. This is based on the voice source parametrization schemes developed both independently and in parallel by Fant and his co-authors (Fant and Lin, 1988; Fant *et al.*, 1994; Fant 1995, 1997) and Alku and Vilkman (Alku and Vilkman, 1996a, 1996b; Alku *et al.*, 2002). In these studies, the application of the ratio between the peak-to-peak amplitude of the glottal flow and the negative peak amplitude of the flow derivative was analyzed in the parametrization of the glottal source. This ratio was shown by Fant and his coauthors to yield "a measure of effective decay time of the glottal flow pulse" and it was initially used as a method to reduce the number of

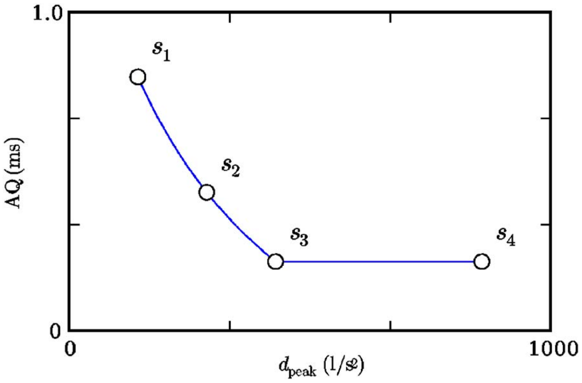

FIG. 3. (Color online) Illusory behavior of amplitude quotient (AQ) as a function of the negative peak amplitude of the flow derivative (d_{peak}); both quantities are expressed on linear scales using arbitrary values on the ordinate and abscissa. Four speech samples (s_1,s_2,s_3,s_4) were created by increasing vocal intensity. Increasing intensity from s_1 to s_3 was carried out by shortening the length of the glottal closing phase, which implies that the value of AQ decreases when d_{peak} is raised. Increasing intensity from s_3 to s_4, however, was achieved by a multiplication of the ac flow of the glottal pulse, which implies that the value of AQ remains constant when d_{peak} is raised.

the LF parameters in modeling the glottal source (Fant *et al.*, 1994; Fant 1995, 1997). The ratio between the two amplitude domain quantities is named the amplitude quotient (AQ) and its normalized version (with respect to the length of the fundamental period) is called the normalized amplitude quotient (NAQ) (Alku *et al.*, 2002). Using the notations given in Fig. 1, the definition of AQ is as follows:[3]

$$\mathrm{AQ} = \frac{f_{ac}}{d_{peak}}. \tag{2}$$

It has been shown that AQ has the following two important features (Alku *et al.*, 2002; Bäckström *et al.*, 2002; Gobl and NiChasaide, 2003). First, AQ is a time-domain quantity that correlates closely with the length of the glottal closing phase, but is always smaller than the true closing phase (except for the nonrealistic triangular-shaped flow, for which AQ becomes equal to the length of the closing phase). Second, computation of AQ is straightforward because it does not involve the extraction of the time instant of glottal closure. Hence, the quantification of the time-based features of the glottal closing phase with AQ (or with NAQ) is more robust against distortion such as the formant ripple present in glottal flows estimated by inverse filtering.

In our previous Studies, AQ (and NAQ) has been used in the quantification of the glottal flow when, for example, the phonation type is changed (Alku and Vilkman, 1996a, b) or as a measure of perceived phonatory pressedness in singing (Sundberg *et al.*, 2004; Björkner *et al.*, in press). Instead of treating AQ as such, the present study introduces a new application for AQ especially well suited for the analysis of intensity regulation. This new means of exploiting AQ is based on a two-dimensional presentation shown in a schematic form in Fig. 3, where AQ (y axis, linear scale) is plotted against d_{peak} (x axis, also on a linear scale). This straightforward presentation will demonstrate how changing d_{peak} which takes place in intensity regulation, will affect AQ in different ranges of d_{peak}.

Let us assume that when an illusory speaker repeats a vowel sound several times, gradually increasing the vocal intensity after each phonation, the corresponding values of d_{peak} and AQ of each speech sample will result as depicted in the two-dimension presentation, shown in Fig. 3. Keeping in mind that SPL has a strong positive correlation with d_{peak} (Gauffin and Sundberg, 1989), it is obvious that the first speech sample, depicted by s_1 in Fig. 3, with the lowest vocal intensity has a small value of d_{peak}. Moreover, it is likely (e.g., Holmberg *et al.*, 1988) that the speaker has used a rounded glottal pulse with gradual opening and closing in the production of this speech sample and, consequently, the length of the glottal closing phase is extended. Given the strong correlation between the length of the glottal closing phase and AQ (Alku *et al.*, 2002), this also implies that the first speech sample has a large value of AQ. The location of the next speech sample, with slightly larger vocal intensity (depicted by s_2 in Fig. 3), is characterized by a larger value of d_{peak} and a smaller value AQ, provided that the intensity increase was achieved by a reduction of the length of the glottal closing phase. Consequently, in this simplified schematic presentation, one would expect that the speech samples of increasing vocal intensity will follow a decreasing function as shown in Fig. 3. According to the definition of AQ [Eq. (2)] the prerequisite for this decrease demands that the increase of d_{peak} [denominator in Eq. (2)] be larger than the increase of f_{ac} [nominator in Eq. (2)]. More specifically, if we also assume that the increase of intensity is achieved by keeping the ac flow constant and by enlarging d_{peak} merely by affecting the shape of the glottal closing phase, the decreasing function between s_1 and s_2 in Fig. 3 will be a hyperbola. [This can be easily seen in Eq. (2), which shows that AQ becomes proportional to $1/d_{peak}$ when f_{ac} is a constant.] Let us continue the simplified example by assuming that after the production of the next sample, depicted by s_3 in Fig. 3, the speaker abruptly changes the intensity regulation strategy so that the length of the glottal closing phase is no longer shortened, but the vocal intensity is still to be raised. In other words, the illusory speaker starts to produce louder speech sounds by amplifying only the amplitude of the flow pulse. Since a (pure) multiplication of a signal with a constant corresponds to multiplying the derivative of the signal with the same constant, it is clear, according to Eq. (2) that the value of AQ will not be affected in this case. In Fig. 3, this implies that the new speech sample, depicted by s_4, has a larger value of d_{peak} than the previous sample, but (exactly) the same value of AQ. In other words, the function between AQ and d_{peak} changes from a hyperbola into a horizontal line between s_3 and s_4.

Even though the above-noted example is illusory, it can be used to analyze the intensity regulation of natural speech. In particular, when speech samples of gradually increasing intensity are expressed in the AQ-d_{peak} presentation, with this scheme it is possible to demonstrate how a speaker's means of increasing his or her vocal intensity changes from affecting the shape of the glottal closing phase to the strategy based on amplifying the flow signal magnitude but keeping the pulse shape more or less constant. In the next section,

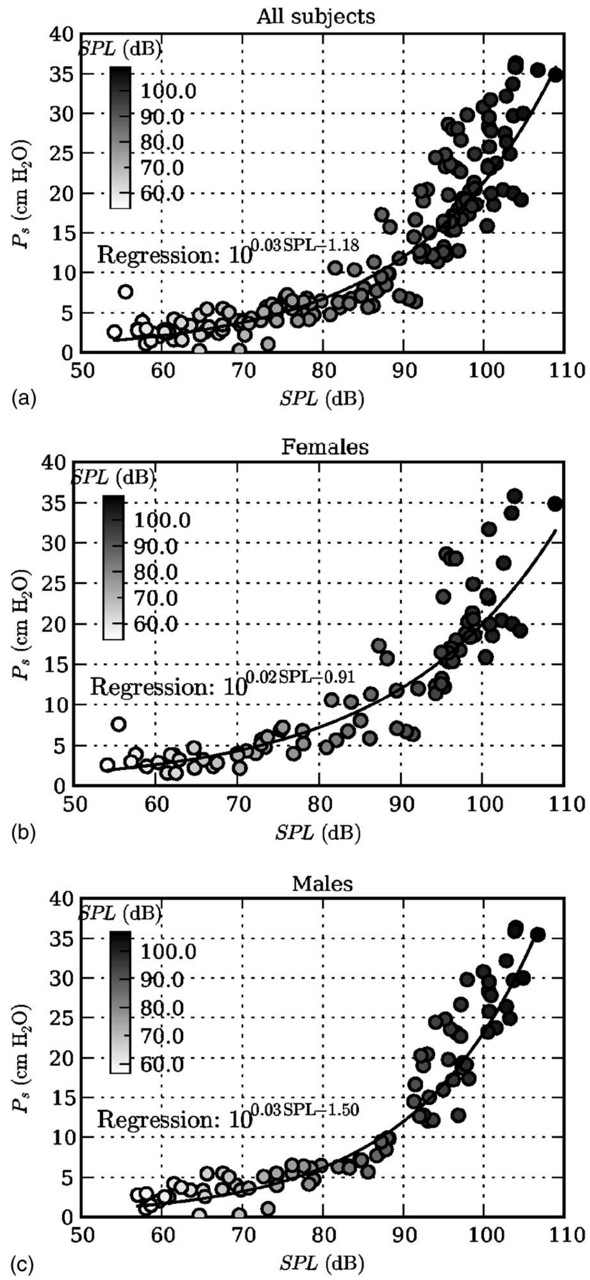

FIG. 4. Subglottal pressure (P_s) as a function of sound pressure level (SPL). Each circle denotes a speech sample whose SPL value is also expressed on a grey scale. Optimal exponential function, fitted by the regression analysis to the data, is shown as a continuous curve. (a) All speakers ($n=10$), (b) female speakers ($n=5$), (c) male speakers ($n=5$).

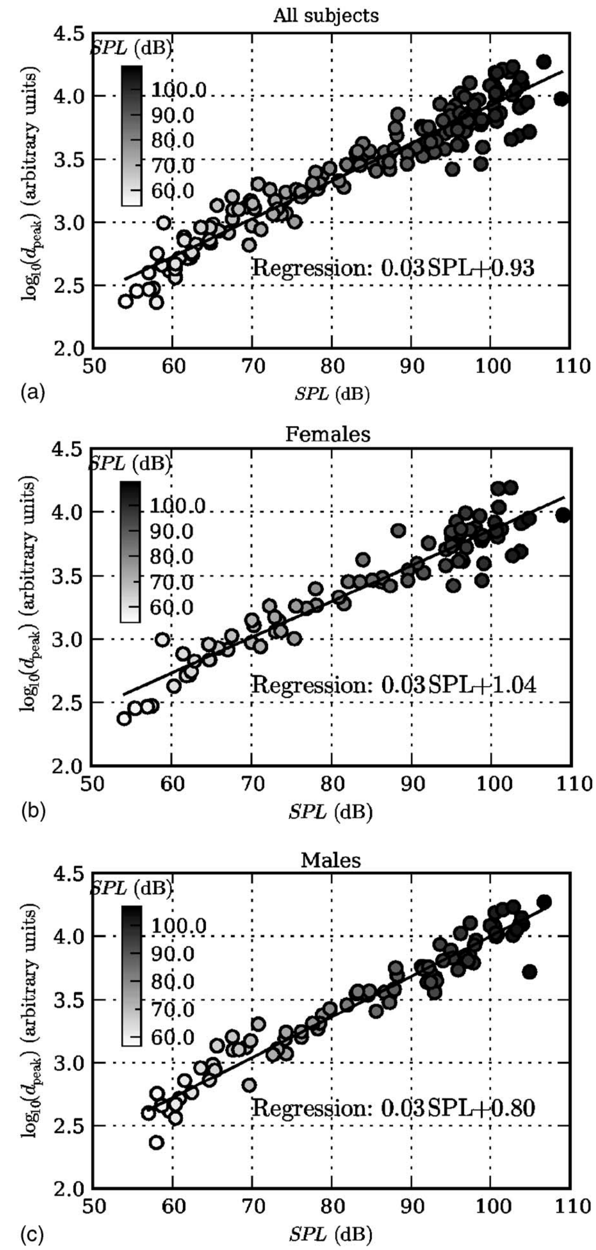

FIG. 5. Negative peak amplitude of the glottal flow derivative (d_{peak}) as a function of SPL. Each circle denotes a speech sample whose SPL value is also expressed on a grey scale. Optimal linear function, fitted by the regression analysis to the data, is shown as a continuous line. (a) All speakers ($n=10$), (b) female speakers ($n=5$), (c) male speakers ($n=5$).

this approach will be used together with other sources of information (subglottal pressure and SPL) extracted from natural speech samples.

III. RESULTS

Since the speaking task in the present study was to create voices of increasing intensity, the data obtained are presented in this section by expressing every sample with a grey scale denoting the SPL value of the sound (see grey scale bar in Figs. 4–6). The data are depicted separately for female and male speakers but also by combining the two genders sexes. The presentation of the results comprises four stages, corresponding to Figs. 4–7, respectively. The reasons for this kind presentation are as follows. First, since subglottal pressure is known to be one of the key parameters underlying intensity regulation (Stevens, 1977), it is reasonable to present the behavior of P_s as a function of SPL which was the parameter that the subjects controlled in the recordings. The data obtained in these analyses are shown in Fig. 4. Second, the negative peak amplitude of the flow derivative is expressed (on a logarithmic scale) as a function of SPL (Fig. 5) in order to demonstrate how the well-known, strong correlation of d_{peak} and SPL (Gauffin and Sundberg, 1989) was reflected in the speech samples of the present study characterized by substantial dynamics in sound pressure level values. Third, the data are presented by using the proposed

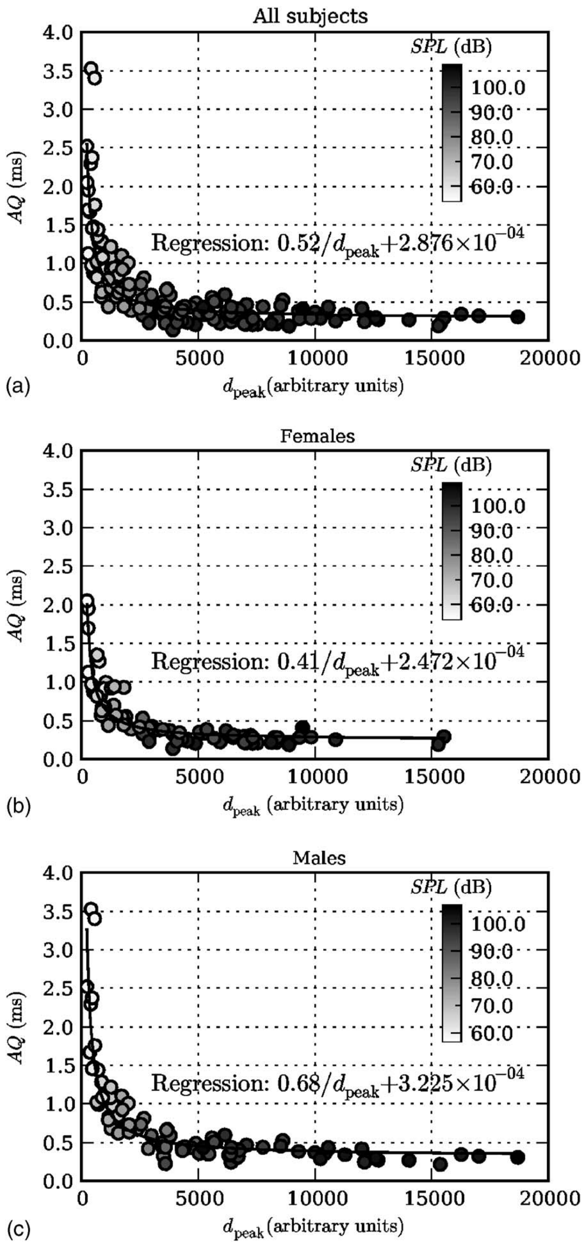

FIG. 6. AQ as a function of the negative peak amplitude of the glottal flow derivative (d_{peak}). AQ is a measure for the glottal closing phase length and it is defined according to Eq. (2) from two amplitude-domain quantities, the peak-to-peak flow, and the negative peak amplitude of the glottal flow derivative. Each circle denotes a speech sample whose SPL value is also expressed on a grey scale. Optimal hyperbolic function, fitted by the regression analysis to the data, is shown as a continuous curve. (a) All speakers ($n=10$), (b) female speakers ($n=5$), (c) male speakers ($n=5$).

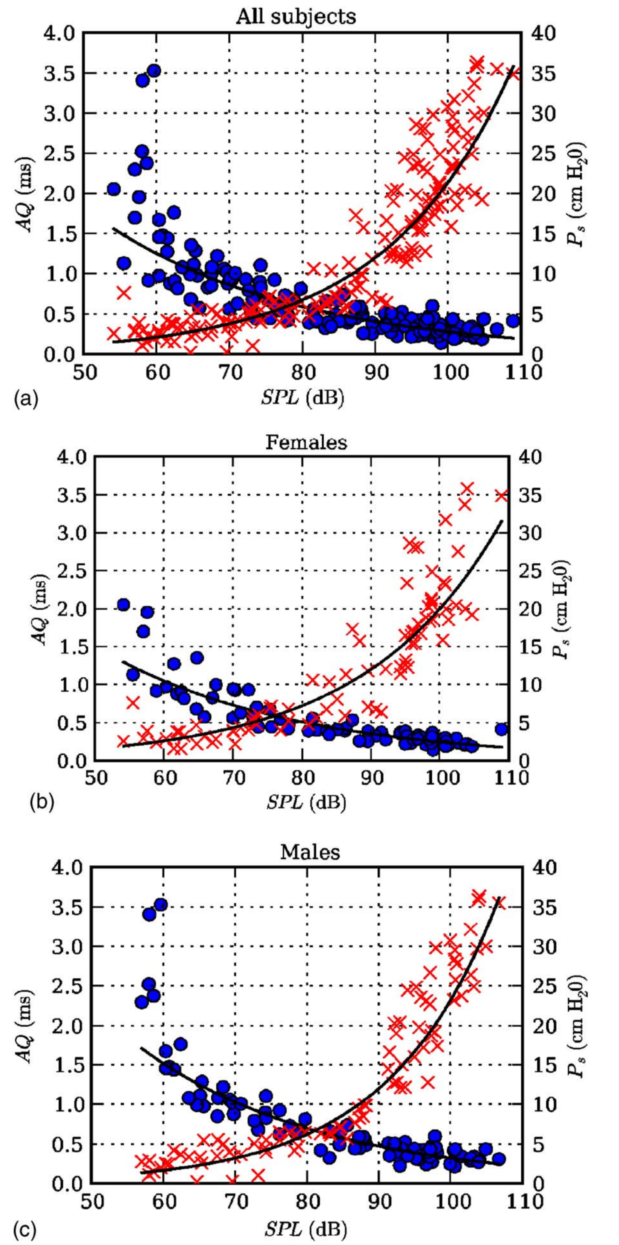

FIG. 7. (Color online) AQ (circles) and subglottal pressure (P_s, crosses) as a function of SPL. Optimal exponential functions, fitted by the regression analysis to AQ and P_s values, are shown by descending and ascending curves, respectively. (a) All speakers ($n=10$), (b) female speakers ($n=5$), (c) male speakers ($n=5$).

AQ-d_{peak} relationship in Fig. 6. Finally, Fig. 7 shows a combinatory expression of both AQ and P_s as a function of SPL in order to demonstrate how these two distinct features underlying intensity regulation behaved as a function of SPL.

The range of SPL values produced by males[4] (57–111 dB) was almost the same as that of females (54–109 dB). As expected, based on previous studies (e.g., Van den Berg, 1956; Ladefoged and McKinney, 1963), subglottal pressure increased when SPL was raised. The lowest pressure values measured equaled 1.6 cm H_2O and 1.1 cm H_2O for male and female subjects, respectively. The largest P_s values of males and females equaled 35.8 cm H_2O and 36.3 cm H_2O, respectively. It is worth noting that the range of P_s values measured in the present study is considerably larger than in many previous studies, in which glottal flow characteristics underlying intensity regulation have been analyzed. The range of P_s was 3.3–15.4 cm H_2O for males and 3.2–13.1 cm H_2O for females in the study by Holmberg *et al.* (1988). In the study by Södersten *et al.* (1995), in which only female voices were analyzed, the range of P_s was 3.2–20.4 cm H_2O.

When the subglottal pressure is plotted against the sound pressure level the data points appear to follow an exponentially rising curve. Figure 4 shows the data scatterplot together with an exponential function that was optimally fit to the data by using regression analysis. In the log-transformed domain, into which the regression line was fitted, the correlation coefficient of SPL and log-transformed P_s was 0.893 (for all subjects), indicating a very large correlation between the subglottal pressure and the sound pressure level.

The strong correlation of SPL and the negative peak of the glottal flow derivative, as demonstrated previously in several studies (e.g., Gauffin and Sundberg, 1989), is supported by the current results as shown by the plot of log-transformed d_{peak} versus SPL depicted in Fig. 5. The correlation of SPL, and log-transformed d_{peak} is also high with a correlation coefficient value of 0.959 for all subjects. The close correlation of d_{peak} and SPL is further emphasized in Fig. 6, which shows the AQ parameter values as a function of d_{peak}: the distribution of the data point value in the grey scale closely follows that of the SPL scale.

Figure 6 illustrates the speech samples using the proposed AQ-d_{peak} graph. The data depicted in Fig. 6 follow the schematic trends discussed, using the hypothetical speaker, in Sec. II D: the value of AQ was largest for the voices of low SPL, but it started to decrease rapidly when the vocal intensity was increased (i.e., when d_{peak} rose). This decline of AQ, however, soon began to approach a saturation point after which there was only a minor change in the parameter value even though d_{peak} rose considerably. This behavior is illustrated in Fig. 6 by the regression curve fit to the data. The constant in the regression formula indicates that while SPL (and therefore, also the value of d_{peak}) was increased, the average value of AQ asymptotically approached 0.2877 ± 0.05290 ms. This value may be considered the average lower limit of AQ in intensity regulation. Furthermore, the curve indicates that at low SPL values, vocal intensity was dominantly controlled by AQ variation.

The different control strategies of vocal intensity for the voices of low and high SPL are further illustrated in Fig. 7, in which both the AQ and P_s values are shown in the same plot. In general, the value of AQ decreased when vocal intensity increased. This behavior is feasible because of two results already shown in previous studies: First, there is a strong correlation between AQ and the length of the glottal closing phase (Alku *et al.*, 2002) and, second, there are many studies in which the shortening of the glottal closing phase coinciding with the raising of the intensity has been reported (e.g., Holmberg *et al.*, 1988). Correspondingly, above approximately 85 dB, P_s changed much more rapidly than AQ, suggesting that in that region the SPL changes were mainly caused by changes in the subglottal pressure. In the intermediate zone, 70–85 dB, both AQ and P_s affected the SPL equally.

The gender sex differences of the above-mentioned curves were also assessed. By plotting subglottal pressures against SPL values for females and males [Figs. 4(b) and 4(c), respectively], the statistical significance of the regression line differences was examined by comparing full and reduced ANCOVA models. The comparison indicated that while there were no significant gender sex differences in the intercept term, the slope did differ significantly ($p=0.0013$). Thus, it is suggested that the subglottal pressure rises more rapidly as a function of SPL for males than for females. When $\log(d_{\mathrm{peak}})$ is plotted against SPL, as shown in Figs. 5(b) and 5(c) for females and males, respectively, the analysis of ANCOVA models indicates that there was a statistically significant difference in the intercept term ($p < 2 \cdot 10^{-16}$), indicating that at a given SPL value, female voices exhibited larger d_{peak} values than males. Even the regression slopes differed significantly ($p=0.0048$), although this effect was not as strong as the intercept. When AQ is plotted against the d_{peak} values [Figs. 6(b) and 6(c) for females and males, respectively], the regression coefficients indicate that the level of the horizontal asymptote of the hyperbola was higher for males, and that the scaling factor of the hyperbola was higher for males than for females as well. The scaling factor value differences are also clearly visible in Figs. 7(b) and 7(c), which show AQ and P_s plotted against the SPL values. In the lower SPL range, AQ values declined more rapidly for males than for females. The significance of the gender sex differences in the hyperbolas was examined by calculating the difference between male data points and the female regression curve, and then by fitting another hyperbola to the acquired cross-residue. The results indicated that there were significant differences both in the k coefficient ($p < 2 \cdot 10^{-16}$) and in the level of the horizontal asymptote c ($p < 2 \cdot 10^{-16}$).

IV. DISCUSSION

Since the publication of the classic studies by Isshiki (Isshiki, 1964, 1965), there have been various investigations on respiratory and laryngeal mechanisms underlying regulation of vocal intensity. Respiratory mechanisms can be examined by either directly measuring the respiratory system or by analyzing respiratory activity using subglottal pressure (Stathopoulos and Sapienza, 1993a). Laryngeal means to control vocal intensity have been typically performed by using an inverse filtering approach in which the estimated glottal flow is quantified in terms of its amplitude-domain features, time-domain features, or both (e.g., Holmberg *et al.*, 1988). Both the respiratory and laryngeal mechanisms have been studied, based mostly on speaking tasks where intensity is varied across a few categories, typically three, namely soft, normal, and loud phonation.

This study proposed a new approach for the analysis of intensity regulation in which a time-domain measure of the glottal flow, the amplitude quotient, is presented as a function of the negative peak amplitude of the flow derivative. The resulting two-dimensional graph, the AQ-d_{peak} presentation, was used in the analysis of the glottal function of voices that varied from extremely soft to very loud. In addition to the proposed AQ-d_{peak} presentation, two conventional measurements, the subglottal pressure and the sound pressure level, were used. The extensive dynamics in both of these measurements, approximately 35 cm H_2O for P_s and 55 dB for SPL, indicate that the speakers took advantage of different mechanisms of intensity regulation when performing the speaking task requested. It is of special interest to note that for the softest voice samples, that is, on the lower SPL range between approximately 55 and 70 dB, the increase of subglottal pressure was very modest even though the SPL values changed by almost 15 dB. This indicates that there must have been other means to vary the vocal intensity. However, for loud speech samples, that is, on the higher SPL range between approximately 95 and 110 dB, P_s rose very rapidly together with the SPL increase, suggesting that for these voice samples, raising the subglottal pressure may become the dominant way to increase the vocal intensity.

In the production of real speech, gradually increasing in intensity, it is unnatural to change abruptly from laryngeal means to respiratory mechanisms. Instead, speakers will most likely start regulating intensity of soft samples mostly by taking advantage of laryngeal means which then will be gradually accompanied by respiratory mechanisms when louder samples are to be produced. In the production of the loudest samples, the regulation is mostly dominated by the respiratory mechanisms. In the present study, this phenomenon was described by the AQ-d_{peak} graph as the convergence of the speech samples of increasing intensity toward an asymptote. This asymptote, available with the help of the proposed presentation, implies that there is a (theoretical) maximum limit of pressedness in intensity regulation. Hence, the presentation of speech samples of different intensity levels on the AQ-d_{peak} graph makes it possible to quantify how control mechanisms underlying regulation of vocal intensity change gradually from laryngeal to respiratory means.

In the study by Holmberg *et al.* (1988), the glottal function was addressed in three intensity categories (soft, normal, and loud) by using a wide range of parameters. The speech data of their study consisted of a short /pæ/ syllable produced in a string of five repetitions. The authors speculated that syllable productions in soft, normal, and loud categories might reflect the use of three different "vocal modes" and therefore they might represent intensity regulation conditions that are not compatible with those used in the production of sustained vowels. The authors of the present study agree with this speculation but, based on the results obtained in this research, we argue that the phonations in distinct "vocal modes" could be explained rather by the use of few intensity categories than by the use of syllable productions. Hence, if intensity regulation is studied over several phonations of gradually increasing SPL values, which was the case in the present study in which every subject produced on average 17 voice samples of different SPL values, it is unlikely that speakers can adapt a special "vocal mode" in each production. Therefore, speech samples analyzed in the present investigation might better reflect a continuous changing of the vocal function from the low intensity voice production to the generation of high intensity speech.

Previous studies on vocal intensity regulation have shown that, as a general trend, the length of the glottal closing phase becomes shorter when intensity rises (e.g., Holmberg *et al.*, 1988). This overall behavior was also corroborated by the present study as indicated by Fig. 7, which shows that the AQ values decrease as a function of SPL. Gender sex differences, however, appeared in the changing AQ when intensity was raised. The present study indicated that the horizontal asymptote level of the AQ-d_{peak} graph was significantly higher for males. In addition, the steepness of the graph was also significantly sharper for males as compared to females. Both of these two observations reflect the physical differences in the vocal fold length between genders sex, which results, in general, in a longer closing phase duration, that is, a larger AQ value, in male glottal pulses than in females. The more abrupt steepness AQ-d_{peak} graphs for males also implies that the male voice production mechanism is able to generate glottal flow pulses in low intensity phonation that have a long closing phase, that is, a large value of AQ, and a small value of d_{peak}. Vocal intensity increase for this kind of glottal pulses can be effectively controlled by laryngeal means resulting in a great decrease in the glottal closing phase. Consequently, the AQ-d_{peak} graph will show a steep decrease in particular for male voices at the lowest SPL range.

Comparing gender sex differences found in the present study to those reported in previous ones is somewhat problematic because of the following issues. First, the present study applied a speaking task in which the intensity was increased gradually by 5 dB steps whereas the majority of the previous investigations have applied speaking tasks comprising only few discrete loudness categories. Second, the extensive dynamic range in the SPL values of the present study implies that the change from low-intensity voice productions to high-intensity phonations might have been accomplished by the use of vocal strategies which are different from those used in studies involving considerably smaller SPL dynamics. Third, and most important, most of the previous studies have used glottal source parametrization methods based on normalized time-based values, such as the closing quotient (ClQ), that is, the ratio between the glottal closing phase duration and the length of the fundamental period (e.g., Holmberg *et al.*, 1988; Sulter and Wit, 1996). The present investigation, however, used the AQ measurement, which quantifies the glottal closing phase in absolute time units. It is, therefore, difficult to gauge how much changes in the glottal closing phase characteristics reported in previous investigations are due to changes in the glottal closing phase length, that is, the numerator of ClQ, and how much they reflect the change of the fundamental period, that is, the denominator of ClQ.

Despite these differences, explaining how the gender sex differences demonstrated by the AQ-d_{peak} of the present study are related to the previous ClQ-based investigations on vocal intensity is feasible. In the study by Holmberg *et al.* (1988), it was found that male voices showed a significantly smaller ClQ value in normal and loud phonation than females. However, there was no significant male-female difference in ClQ for soft voices. We may speculate that these results reported by the above-mentioned study on ClQ are actually in line with the results of the present study even though the relationship between the two is perhaps not the most obvious one. It is, for example, possible that in producing soft voices both the female and male speakers in that study might have used the same "vocal mode" resulting in glottal pulses that are of similar shapes, but of different time-lengths, for the two genders sexes. [A stylized illustration is provided in Fig. 10, p. 519, in Holmberg *et al.* (1988)]. Parametrization of this kind of glottal pulse will in fact result in small ClQ differences between males and females, as reported in Holmberg *et al.* (1988), but, importantly, in larger

AQ values in male voices than in female speech as, indeed, was indicated by the AQ-d_{peak} graphs of low-intensity voices in the present study. Moreover, it can be argued that the speakers in the study by Holmberg *et al.* (1988) most likely used means other than respiratory ones when changing vocal intensity from soft to normal speech, because if this intensity change was based solely on the respiratory mechanism, the same "vocal mode" would have been used by both males and females in normal loudness as well. Consequently, one would not have expected to find a gender sex difference in ClQ in normal loudness either. However, the existence of the gender sex difference in normal and loud phonations in the study by Holmberg *et al.* (1988) indicates that the intensity increase from soft voices was accomplished by means other than respiratory; that is, most likely by changes in the time-based characteristics of the glottal closing phase. In the present study, this indicates the occurrence of voice samples on a steep decreasing portion of the AQ-d_{peak} graphs when intensity is increased at the lowest part of the SPL range. Finally, a possible explanation why ClQ values of males in the study by Holmberg *et al.* (1988) were *smaller* than those of females in normal and loud speech while the horizontal level of the AQ-d_{peak} graphs of males in the present study were *larger* than that of females is the difference in the normalization of the time-based parameters involved. In other words, it is possible that if intensity increase in female and male voices is accomplished by the reduction of both the glottal closing phase and the fundamental period, but if the latter happens more rapidly for males, a smaller ClQ value is to be expected for them, as reported in Holmberg *et al.* (1988). However, since the absolute length of the closing phase is still typically larger in male speech, it is understandable why the proposed AQ-d_{peak} graph showed a higher horizontal asymptote level for males than for females.

V. CONCLUSIONS

Regulation of vocal intensity was analyzed by means of a new method, the AQ-d_{peak} graph, from voices whose SPL varied from extremely low to very high. When vocal intensity was increased, the speech samples first showed a rapidly decreasing trend when expressed on the proposed AQ-d_{peak} graph. The hypothetical correspondence of this behavior is a hyperbola, indicating that the speakers have mainly manipulated the length of the glottal closing phase in the regulation of vocal intensity. When intensity was further raised, the location of the speech samples on the AQ-d_{peak} converged toward a horizontal line, the asymptote of the hyperbola. This indicates that the decrease of AQ starts to diminish even though d_{peak} continues to rise which, in turn, implies that the speakers have started to increase their vocal intensity by amplifying the ac flow of the glottal pulse.

When interpreted according to Isshiki (1964, 1965), the behavior of the AQ-d_{peak} graph, as summarized earlier, implies that the intensity regulation strategy changes from laryngeal means, represented by the rapidly decreasing part of the hyperbola, to respiratory mechanisms, represented by the horizontal asymptote. The behavior of the AQ-d_{peak} graph in the speech samples of the current study was in line with results reported in several other investigations on intensity regulation in showing that increasing vocal intensity corresponds typically to the decreasing of the glottal closing phase, reflected in the current study by AQ, and to the increasing of the negative peak amplitude of the glottal flow derivative. In contrast to many previous studies, however, the current research proposes a new means to visualize the function of voice production, which might be of use in investigating intensity regulation of speech and could supplement existing voice production analysis tools such as the phonetogram. The proposed presentation is easy to implement because the only information needed is the glottal flow wave form estimated by inverse-filtering the acoustic speech pressure signal.

ACKNOWLEDGMENTS

This work was supported by the Academy of Finland (Project No. 200859), by the European Community's Human Potential Programme under Contract No. HPRN-CT-2002-00276 [HOARSE-network], and by the KAUTE foundation.

[1]Instead of using the term "subglottal pressure" Titze frequently uses the term "lung pressure" to refer to the mean subglottal pressure as estimated during oral occlusion (Fant, 1997).

[2]The energy of a digital signal is computed as the (infinite) sum of squares of the signal samples (Oppenheim and Schafer, 1975).

[3]It should be noticed in Eq. (2) that the domain of AQ is time, because this quotient is defined as a ratio between a flow value and a value of the *time derivative* of the flow. If inverse filtering is based on digital signal processing, which is typical in voice source analysis today, the values of f_{ac} and d_{peak} usually extracted from discrete-time wave forms that are expressed using integer numbers as the time variable. In this case, AQ in Eq. (2) needs to be divided by the sampling frequency in order to express the parameter value in seconds.

[4]Measurement of subglottal pressure failed for some of the high SPL speech samples produced by male subjects due to clipping of the intraoral pressure waveforms in the analog-to-digital conversion. Therefore, these phonations could not be included in those analyses in which subglottal pressure values are required, but they were used in all other analyses.

Åkerlund, L., Gramming, P., and Sundberg, J. (**1992**). "Phonetogram and averages of sound pressure levels and fundamental frequencies of speech: Comparison between female singers and nonsingers," J. Voice **6**, 55–63.

Alku, P. (**1992**). "Glottal wave analysis with Pitch Synchronous Iterative Adaptive Inverse Filtering," Speech Commun. **11**, 109–118.

Alku, P., Bäckström, T., and Vilkman, E. (**2002**). "Normalized amplitude quotient for parameterization of the glottal flow," J. Acoust. Soc. Am. **112**, 701–710.

Alku, P., and Vilkman, E. (**1994**). "Estimation of the glottal pulseform based on discrete all-pole modeling," in Proceedings of the International Conference on Spoken Language Processing, Yokohama, Japan, pp. 1619–1622.

Alku, P., and Vilkman, E. (**1996a**). "A comparison of glottal voice source quantification parameters in breathy, normal, and pressed phonation of female and male speakers," Folia Phoniatr Logop **48**, 240–254.

Alku, P., and Vilkman, E. (**1996b**). "Amplitude domain quotient for characterization of the glottal volume velocity waveform estimated by inverse filtering," Speech Commun. **18**, 131–138.

Alku, P., Vilkman, E., and Laukkanen, A.-M. (**1998**). "Estimation of amplitude features of the glottal flow by inverse filtering speech pressure signals," Speech Commun. **24**, 123–132.

Alku, P., Vintturi, J., and Vilkman, E. (**1999**). "On the linearity of the relationship between the sound pressure level and the negative peak amplitude of the differentiated glottal flow in vowel production," Speech Commun. **28**, 269–281.

Alku, P., Vintturi, J., and Vilkman, E. (**2002**). "Measuring the effect of fundamental frequency raising as a strategy for increasing vocal intensity in soft, normal and loud phonation," Speech Commun. **38**, 321–334.

Björkner, E., Sundberg, J., and Alku, P. (in press). "Subglottal pressure and normalized amplitude quotient variation in classically trained baritone singers," Logoped. Phoniatr. Vocol.
Bouhuys, A., Mead, J., Proctor, D., and Stevens, K. (**1968**). "Pressure-flow events during singing," Ann. N.Y. Acad. Sci. **155**, 165–176.
Bäckström, T., Alku, P., and Vilkman, E. (**2002**). "Time-domain parameterization of the closing phase of glottal airflow waveform from voices over a large intensity range," IEEE Trans. Speech Audio Process. **10**, 186–192.
Coleman, R. F. (**1993**). "Sources of variation in phonetograms," J. Voice **7**, 1–14.
Coleman, R. F., Mabis, J. H., and Hinson, J. K. (**1977**). "Fundamental frequency-sound pressure level profiles of adult male and female voices," J. Speech Hear. Res. **20**, 197–204.
Cummings, K. E., and Clements, M. A. (**1995**). "Analysis of the glottal excitation of emotionally styled and stressed speech," J. Acoust. Soc. Am. **98**, 88–98.
Damste, H. (**1970**). "The phonetogram," Pract. Otorhinolaryngol. (Basel) **32**, 185–187.
Dromey, C., Stathopoulos, E., and Sapienza, C. (**1992**). "Glottal airflow and electroglottographic measures of vocal function at multiple intensities," J. Voice **6**, 44–54.
El-Jaroudi, A., and Makhoul, J. (**1991**). "Discrete all-pole modeling," IEEE Trans. Acoust., Speech, Signal Process. **39**, 411–423.
Fant, G. (1982). "Preliminaries to analysis of the human voice source," STL-QPSR 4, Dept. of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, pp. 1–27.
Fant, G. (**1993**). "Some problems in voice source analysis," Speech Commun. **13**, 7–22.
Fant, G. (1995). "The LF-model revisited. Transformations and frequency domain analysis," STL-QPSR 2-3, Dept. of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, pp. 119–156.
Fant, G. (**1997**). "The voice source in connected speech," Speech Commun. **22**, 125–139.
Fant, G., Kruckenberg, A., Liljencrants, J., and Båvegård, M. (1994). "Voice source parameters in continuous speech. Transformation of LF-parameters," in Proceedings of the International Conference on Spoken Language Processing, Yokohama, Japan, pp. 1451–1454.
Fant, G., and Lin, Q. (1988). "Frequency domain interpretation and derivation of glottal flow parameters," STL-QPSR 2-3, Dept. of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm, pp. 1–21.
Finnegan, E., Luschei, E., and Hoffman, H. (**2000**). "Modulations in respiratory and laryngeal activity associated with changes in vocal intensity during speech," J. Speech Lang. Hear. Res. **43**, 934–950.
Gauffin, J., and Sundberg, J. (**1989**). "Spectral correlates of glottal voice source waveform characteristics," J. Speech Hear. Res. **32**, 556–565.
Gobl, C., and Ni Chasaide, A. (**2003**). "Amplitude-based voice source parameters for measuring voice quality," in Proceedings of the ISCA Workshop on Voice Quality: Functions, Analysis and Synthesis, Geneva, Switzerland, pp. 151–156.
Gramming, P., and Sundberg, J. (**1988**). "Spectrum factors relevant to phonetogram measurement," J. Acoust. Soc. Am. **83**, 2352–2360.
Gramming, P., Sundberg, J., Ternstöm, S., Leanderson, R., and Perkins, W. (**1988**). "Relationship between changes in voice pitch and loudness," J. Voice **2**, 118–126.
Hertegård, S., and Gauffin, J. (**1991**). "Insufficient vocal fold closure as studied by inverse filtering," in *Vocal Fold Physiology: Acoustics, Perceptual and Physiological Aspects of Voice Mechanisms*, edited by J. Gauffin and B. Hammarberg (Singular, San Diego), pp. 243–250.
Hertegård, S., Gauffin, J., and Lindestad, P-Å. (**1995**). "A comparison of subglottal and intraoral pressure measurements during phonation," J. Voice **9**, 149–155.
Hirano, M., Vennard, W., and Ohala, J. (**1970**). "Regulation of register, pitch and intensity of voice," Folia Phoniatr (Basel) **22**, 1–20.
Hodge, S., Colton, R., and Kelley, R. (**2001**). "Vocal intensity characteristics in normal and elderly speakers," J. Voice **7**, 503–511.
Holmberg, E., Hillman, R., and Perkell, J. (**1988**). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," J. Acoust. Soc. Am. **84**, 511–529.
Isshiki, N. (**1964**). "Regulatory mechanism of voice intensity variation," J. Speech Hear. Res. **7**, 17–29.
Isshiki, N. (**1965**). "Vocal intensity and air flow rate," Folia Phoniatr (Basel) **17**, 92–104.
Ladefoged, P., and McKinney, N. P. (**1963**). "Loudness, sound pressure, and subglottal pressure in speech," J. Acoust. Soc. Am. **35**, 454–460.
Löfqvist, A., Carlborg, B., and Kitzing, P. (**1982**). "Initial validation of an indirect measure of subglottal pressure during vowels," J. Acoust. Soc. Am. **72**, 633–635.
Makiyama, K., Yoshihashi, H., Mogitate, M., and Kida, A. (**2005**). "The role of adjustment of expiratory effort in the control of vocal intensity: Clinical assessment of phonatory function," Otolaryngol.-Head Neck Surg. **132**, 641–646.
Oppenheim, A., and Schafer, R. (**1975**). *Digital Signal Processing* (Prentice Hall, Englewood Cliffs, NJ).
Sapienza, C., and Stathopoulos, E. (**1994**). "Comparison of maximum flow declination rate: Children versus adults," J. Voice **8**, 240–247.
Sapienza, C., Stathopoulos, E., and Dromey, C. (**1998**). "Approximations of open quotient and speed quotient from glottal airflow and EGG waveforms: Effects of measurement criteria and sound pressure level," J. Voice **12**, 31–43.
Stathopoulos, E., and Sapienza, C. (**1993a**). "Respiratory and laryngeal function of women and men during vocal intensity variation," J. Speech Hear. Res. **36**, 64–75.
Stathopoulos, E., and Sapienza, C. (**1993b**). "Respiratory and laryngeal measures of children during vocal intensity variation," J. Acoust. Soc. Am. **94**, 2531–2543.
Stevens, K. N. (**1977**). "Physics of laryngeal behavior and larynx models," Phonetica **34**, 264–279.
Strik, H., and Boves, L. (**1992**). "Control of fundamental frequency, intensity and voice quality in speech," J. Phonetics **20**, 15–25.
Sulter, A., and Wit, H. (**1996**). "Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age," J. Acoust. Soc. Am. **100**, 3360–3373.
Sundberg, J., Fahlstedt, E., and Morell, A. (**2005**). "Effects on the glottal voice source of vocal loudness variation in untrained female and male voices," J. Acoust. Soc. Am. **117**, 879–885.
Sundberg, J., Thalen, M., Alku, P., and Vilkman, E. (**2004**). "Measuring perceived phonatory pressedness in singing from flow glottograms," J. Voice **18**, 56–62.
Sundberg, J., Titze, I., and Scherer, R. (**1993**). "Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source," J. Voice **7**, 15–29.
Södersten, M., Hertegård, S., and Hammarberg, B. (**1995**). "Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women," J. Voice **9**, 182–197.
Titze, I. (**1992a**). "Phonation threshold pressure: A missing link in glottal aerodynamics," J. Acoust. Soc. Am. **91**, 2926–2935.
Titze, I. (**1992b**). "Acoustic interpretation of the voice range profile (Phonetogram)," J. Speech Hear. Res. **35**, 21–43.
Titze, I. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).
Titze, I., and Sundberg, J. (**1992**). "Vocal intensity in speakers and singers," J. Acoust. Soc. Am. **91**, 2936–2946.
Van den Berg, J. W. (**1956**). "Direct and indirect determination of the mean subglottic pressure," Folia Phoniatr. **8**, 1–24.
Vilkman, E., Lauri, E-R., Alku, P., Sala, E., and Sihvo, M. (**1997**). "Loading-changes in time-based parameters of glottal flow waveforms in different ergonomic conditions," Folia Phoniatr. **49**, 247–263.
Young, H. D., and Freedman, R. A. (**1996**). *University Physics* (Addison-Wesley, Reading, MA), p. 612.