Just Noticeable Differences of Open Quotient and Asymmetry Coefficient in Singing Voice

*†Nathalie Henrich, *Gunilla Sundin, *Daniel Ambroise, †Christophe d'Alessandro, *Michèle Castellengo, and †Boris Doval

Paris and Orsay, France

Summary: This study aims to explore the perceptual relevance of the variations of glottal flow parameters and to what extent a small variation can be detected. Just Noticeable Differences (JNDs) have been measured for three values of open quotient (0.4, 0.6, and 0.8) and two values of asymmetry coefficient (2/3 and 0.8), and the effect of changes of vowel, pitch, vibrato, and amplitude parameters has been tested. Two main groups of subjects have been analyzed: a group of 20 untrained subjects and a group of 10 trained subjects. The results show that the JND for open quotient is highly dependent on the target value: an increase of the JND is noticed when the open quotient target value is increased. The relative JND is constant: $\Delta O_q/O_q = 14\%$ for the untrained and 10% for the trained. In the same way, the JND for asymmetry coefficient is also slightly dependent on the target value–an increase of the asymmetry coefficient value leads to a decrease of the JND. The results show that there is no effect from the selected vowel or frequency (two values have been tested), but that the addition of a vibrato has a small effect on the JND of open quotient. The choice of an amplitude parameter also has a great effect on the JND of open quotient.

Key Words: Singing voice synthesis—Glottal flow—Open quotient—Asymmetry coefficient—Just Noticeable Differences.

INTRODUCTION

It is now well known that the voice source affects the perceived voice quality. Many studies, such as the ones done by Monsen and Engebretson,[1] Holmberg et al,[2] or Gauffin and Sundberg,[3] show that the temporal aspect of glottal flow varies with intensity and timbre changes. Sundberg et al[4] illustrated three ways in which the sound pressure level (SPL) can be increased at a fixed fundamental frequency, in the case of male singing: increase of the maximum amplitude of flow passing through the glottis, modification of the glottal flow asymmetry, or increase of the negative peak amplitude of the differentiated glottal flow. The open quotient is another glottal flow parameter that has been proved to vary with modification of vocal intensity.[2,5–8] Because a correlation between these glottal flow parameters and vocal intensity has been highlighted in these studies, the question can be raised as to whether a measured

Accepted for publication July 23, 2002.

From the *LAM (UPMC, CNRS, Ministère de la culture), Paris; †LIMSI-CNRS, Orsay, France.

Address correspondence and reprint requests to Nathalie Henrich, Laboratoire d'Acoustique Musicale, 11 rue de Lourmel, 75015 Paris, France.
E-mail: nathalie.henrich@lam.jussieu.fr

Journal of Voice, Vol. 17, No. 4, pp. 481–494

0892-1997/2003 $30.00+0
doi:10.1067/S0892-1997(03)00005-5

difference in such glottal flow parameters can be perceived or not. One approach to this problem consists in predicting the perceptual relevance of spectral changes by applying auditory models and measuring perceptual distances.[9–12] This approach is completed experimentally, using psychoacoustic tests to determine the Just Noticeable Differences (JNDs). A JND can be defined as a perceptual measure of the extent to which a parameter can be varied without being perceived. Such measure can be used to assess the perceptual validity of a parameter variation. One study, conducted by Scherer et al,[13] has been dedicated to measuring JNDs for normal glottal source characteristics. The study focused on two parameters: the open quotient (O_q), defined as the ratio between the open time and the fundamental period, and the speed quotient (S_q), defined as the ratio between the flow rise time (from baseline to peak flow) and the flow fall time (from peak flow back to baseline). They determined the JND around a single value for each parameter ($O_q = 0.6$ and $S_q = 2$), with $f_0 = 130.39$ Hz. The synthetic stimuli were chosen to be either the glottal flow or the acoustic signal obtained after vocal tract filtering in the case of the vowel [AA]. For the glottal source, the results showed a JND of 0.0264 for open quotient and 0.154 for speed quotient. For the vowel signal, the JNDs were 0.0344 for O_q and 0.319 for S_q.

The current study is a sequel to that of Scherer et al.[13] The previous study was limited to a single value of the glottal flow parameter, one frequency and one vowel. The main goal of this study is to investigate whether the JNDs would be modified when the glottal flow parameter value is changed. In addition, will the choice of frequency or vowel have an effect on the JNDs? We will focus on open quotient and asymmetry coefficient. The aymmetry coefficient (α_m) is a dimensionless glottal flow parameter defined as the ratio between the flow rise time and the open time. It is equivalent to the speed quotient, as shown by the following relationship: $\alpha_m = \frac{S_q}{1 + S_q}$. This parameter is represented in Figure 1(a), in addition to open quotient and other glottal flow parameters, and the relationship between speed quotient and asymmetry coefficient is plotted in Figure 1(b). This parameter was introduced by Doral and d'Alessandro[14] to simplify the equations of glottal flow models. Another advantage of using the asymmetry coefficient instead of the speed quotient is that the values of this parameter are more easily understandable: α_m ranges between 0 and 1 (which corresponds to $0 < S_q < +\infty$), with typical values between 0.5 ($S_q = 1$) and 0.8 ($S_q = 4$). For $\alpha_m < 0.5$ ($S_q < 1$), the glottal pulse is skewed to the left, for $\alpha_m = 0.5$, the glottal pulse is symmetric, and for $\alpha_m > 0.5$, the glottal pulse is skewed to the right.

METHOD

Subjects

A total of 27 subjects participated in the test (22 males and 5 females, 23 to 58 years of age, with median age of 27 years). As Scherer et al[13] found no statistically significant difference between genders, the gender will not be taken into account here. All but three subjects had some degree of listening training as they were musicians and/or acousticians. The listening test lasted about one and a half hours. Prior to the test, each participant was subjected to a hearing threshold screening for each ear at 20 dB, for frequencies ranging from 250 Hz to 8000 Hz (250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000, and 8000 Hz). Seven subjects did not pass the screening at high frequencies; nevertheless, they participated in the listening test, but their results will be discussed apart from the results of the 20 other subjects who passed the screening. Among the 20 subjects who passed the screening, 10 accepted to repeat the listening test 3 times, within a span of several weeks.

Synthesis of stimuli

In the Scherer et al study, the synthesis program SPEAK[15] was used for creating the stimuli, which were both the glottal flow and the output pressure signal "at the lips." The fundamental frequency was kept fixed ($f_0 = 130.39$ Hz). In this study, the naturalness of the synthesis was our priority, so as to provide stimuli that would be close to real sounds. Thus, we chose not to keep the fundamental frequency constant, and we chose to use only the output pressure as a stimulus. For this purpose, we have developed a method for high-quality singing synthesis of vowels, where the fundamental frequency

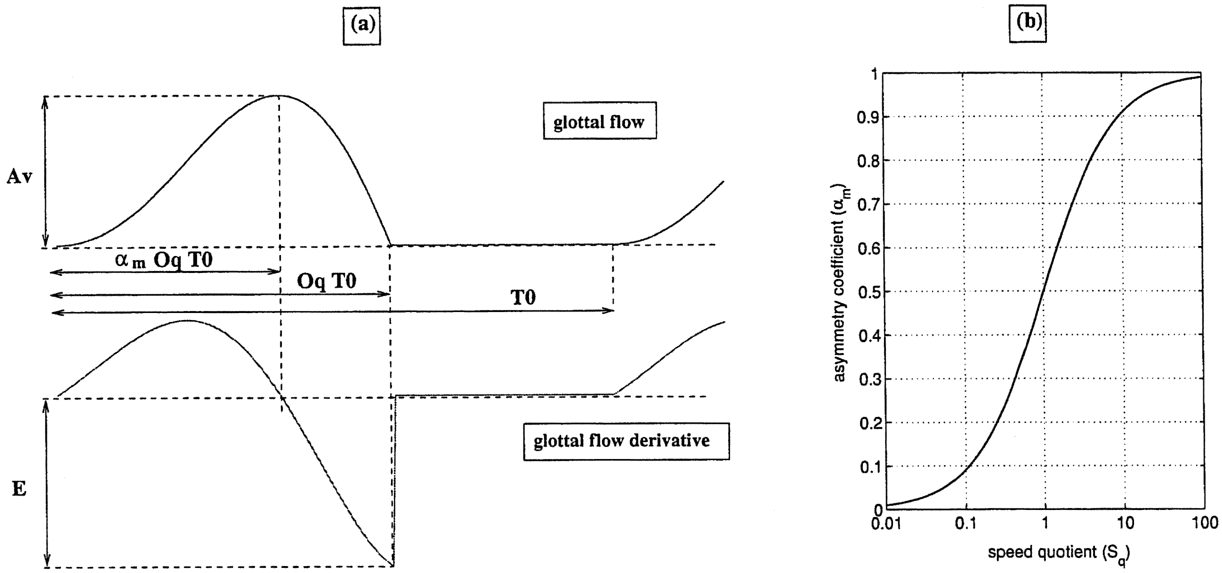

FIGURE 1. Panel (**a**): Glottal flow parameters represented on one period of the glottal flow and its derivative. A_v, amplitude of voicing, E, negative peak amplitude of the differentiated glottal flow, T_0, fundamental period, O_q, open quotient, α_m, asymmetry coefficient. Panel (**b**). Relationship between the speed quotient (logarithmic scale) and the asymmetry coefficient.

is varied to replicate natural vibrato. The method can be described as follows:

(1) Detection of glottal closing instants.

A glottal closing instant (GCI) can be defined on the differentiated glottal flow as the instant of the maximum negative peak amplitude. The glottal closing instants are detected on the derivative of an electroglottographic (EGG) signal using the wavelet transform method developed by Vu Ngoc and d'Alessandro.[16,17] The EGG signal was recorded during sustained vowels sung by a baritone singer. A database of EGG signals was collected by Henrich[18] and the GCIs were extracted from this set.

(2) Synthesis of the glottal flow derivative waveform

A glottal flow derivative impulse is synthesized between two following glottal closing instants (pitch-synchronous synthesis) using the LF model[19] with a set of fixed glottal flow parameters.

(3) Synthesis of the voiced signal

The synthesized glottal flow derivative signal is then filtered by a fixed transfer function that has been estimated by linear prediction of a real signal.

The use of GCIs extracted from a real signal seems key to a natural-sounding synthesis. Indeed, it provides small time variation information, to which our ears may be very sensitive.

Two pitches have been selected for this study: C3 (f_0 around 130 Hz) and G3 (f_0 around 196 Hz). They corresponded to a frequency region where the baritone singer felt very comfortable. The first pitch was chosen to be similar to the previous study for allowing comparisons and the second one was chosen a fifth above. The corresponding f_0 curves, deduced from the GCI by $f_0(i) = 1/(T_{c(i)} - T_{c(i-1)})$, are shown in Figure 2. Two vowels (french [AA] and [IY]*) have also been selected, one with a relative high first formant frequency ([AA], F_1 = 560 Hz as measured on this baritone singer) and the other one with a low first formant frequency ([IY], F_1 = 330 Hz). Figure 3 shows the corresponding transfer functions, which were estimated by linear prediction of a sustained vowel sung by the baritone.

The parameters used to describe the glottal flow impulse are represented in Figure 1(a). Using this

*The vowels are represented in the ARPABET phonetic alphabet.[20]

set of glottal flow parameters, the equations for the LF model have been rewritten as follows, in the case of abrupt closure (no return phase). The fundamental period is noted as:

$$T_0 = \frac{1}{f_0} = T_{c(i)} - T_{c(i-1)}$$

If we assume that the time $t = 0$ corresponds to a glottal opening instant, a one-period glottal flow u_g, and its derivative u'_g can be expressed as (Doval and d'Alessandro[14])

for $0 < t < O_q T_0$,

$$u_g(t) = A_v n_g \left(\frac{t}{O_q T_0}, \alpha_m \right) \tag{2}$$

$$u'_g(t) = \frac{A_v}{O_q T_0} n'_g \left(\frac{t}{O_q T_0}, \alpha_m \right)$$

for $O_q T_0 < t < T_0$ (no return phase),

$$u_g(t) = 0$$

$$u'_g(t) = 0$$

n_g is a normalized function of the glottal flow, ranging between 0 (baseline) and 1 (peak flow). n'_g corresponds to its derivative. Both functions depend only on the asymmetry coefficient parameter. For the LF model, these functions are given by

$$n_g(t, \alpha_m) = \frac{1 + e^{at} \left(a \frac{\alpha_m}{\pi} \sin\left(\frac{\pi}{\alpha_m} t \right) - \cos\left(\frac{\pi}{\alpha_m} t \right) \right)}{1 + e^{a\alpha_m}} \tag{3}$$

$$n'_g(t, \alpha_m) = \frac{e^{at} \left(a^2 + \left(\frac{\pi}{\alpha_m} \right)^2 \right) \sin\left(\frac{\pi}{\alpha_m} t \right)}{\frac{\pi}{\alpha_m} (1 + e^{a\alpha_m})}$$

with a being the solution of the implicit equation:
$e^{-a} + a \frac{\alpha_m}{\pi} \sin\left(\frac{\pi}{\alpha_m} \right) - \cos\left(\frac{\pi}{\alpha_m} \right) = 0.$

The amplitude parameter, A_v, can easily be replaced by the negative peak amplitude of the differentiated glottal flow, E, and the relation between both parameters is given by:

$$E = \frac{A_v}{O_q T_0} \frac{-e^a \sin\left(\frac{\pi}{\alpha_m} \right) \left(a^2 + \left(\frac{\pi}{\alpha_m} \right)^2 \right)}{\frac{\pi}{\alpha_m} (1 + e^{a\alpha_m})} \tag{4}$$

Default values used in this study are: $A_v = 1$, $O_q = 0.6$, $\alpha_m = 2/3$.

Using analytical formulations of these equations in the frequency domain, we can discuss the spectral correlates of a variation of open quotient, O_q, or asymmetry coefficient, α_m.[14,18,21,22] Figures 4 and 5 show the effects of a variation of O_q and α_m in time and frequency domains. The spectral reinforcement that can be seen in this figure refers to the so-called "glottal formant[23]."[†] In the lower part of the spectrum, the open quotient modifies the glottal formant frequency, whereas the asymmetry coefficient modifies mainly the glottal formant bandwidth. In the upper part of the spectrum, both parameters affect the spectral slope amplitude as this amplitude is related to the parameter E. Indeed, a variation of O_q or α_m implies a variation of E when the amplitude of voicing, A_v, is kept constant.

Methodology

Each listening test consisted of nine sessions in which only one parameter was varied. It was preceded by a training session during which the subject was acquainted with the task, on a three-run example. Table 1 gives the details of the nine sessions. Three sessions were designed to the determination of JNDs around 3 values of open quotient: $O_q = 0.4$ (session 1), $O_q = 0.6$ (session 3), and $O_q = 0.8$ (session 6). Two other sessions were designed to the determination of JNDs around 2 values of asymmetry coefficient: $\alpha_m = 2/3$ (session 7) and $\alpha_m = 0.8$ (session 8). In addition, four sessions were designed to explore whether a JND of open quotient (around $O_q = 0.6$) would be dependent on the choice of the following parameters, as compared to a reference session (session 3):

[†]As the term "glottal formant" may be confusing, we would like to define it. This term is used to designate a spectral maximum in the glottal source spectrum. The glottal formant is not related to any resonance phenomenon.

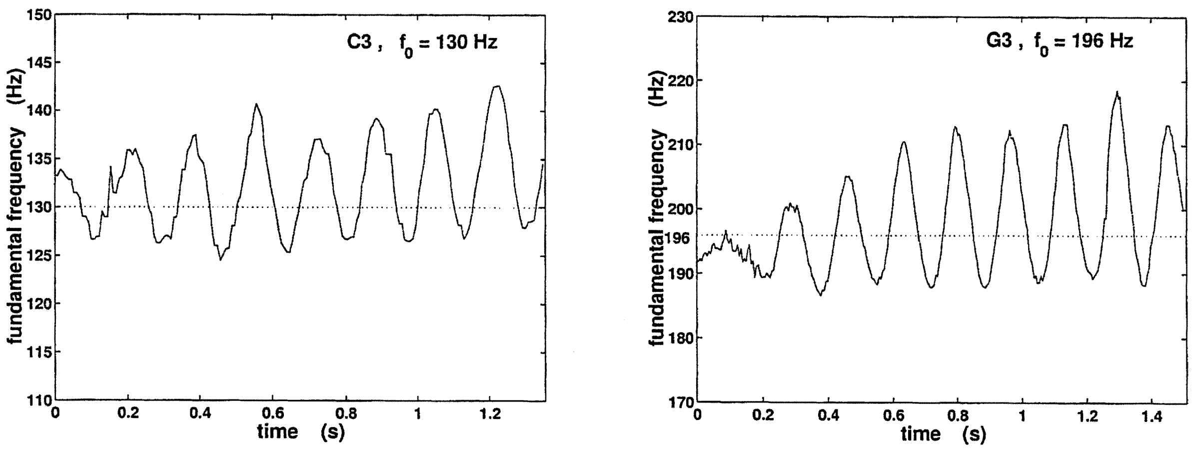

FIGURE 2. Plot of fundamental frequency, calculated from the GCI (T_{GCI}) detected in the case of sustained vowels sung by a baritone on two pitches C3 and G3.

- The fundamental frequency: f_0 = 130 Hz (session 5), as compared to f_0 = 196 Hz.
- The vowel: Vowel [IY] (session 9), as compared to vowel [AA].
- The vibrato: For improving the synthesis quality, the fundamental frequency is varied to replicate a natural vibrato. What would be the effect on the JND measure of an absence of vibrato (fixed fundamental frequency, session 2)?
- The amplitude parameter: In all but one session, the amplitude of voicing is kept constant, which implies that the negative peak amplitude of the differentiated glottal flow (E) varies according to Equation 4. Would the JND measure be modified if this parameter is kept constant, which implies a variation of the amplitude of voicing (session 4)?

The psychoacoustic method for measuring the JND is a three-interval forced choice two-down one-up adaptative procedure.[13,24] It has been implemented in Matlab and adapted to the LISE psychoacoustic test interface.[25] Each of the nine sessions were completed in 40–90 trials. At each trial, the subject listened to three sequential sounds, separated by a 500 ms pause. Two of them were the standard stimulus, and the third was the comparison stimulus, which was varied during the session depending on the subject's answers. The subject was

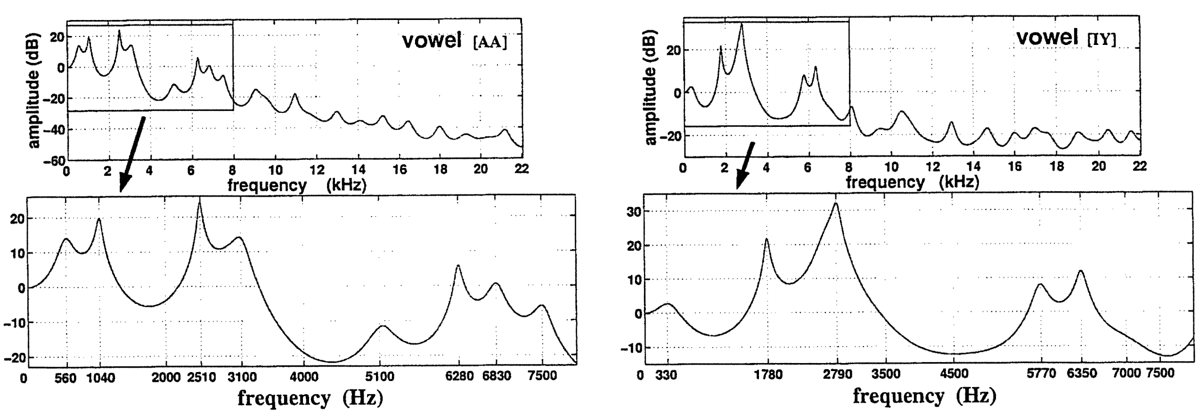

FIGURE 3. Vocal tract transfer function in the case of vowel [AA] and vowel [IY].

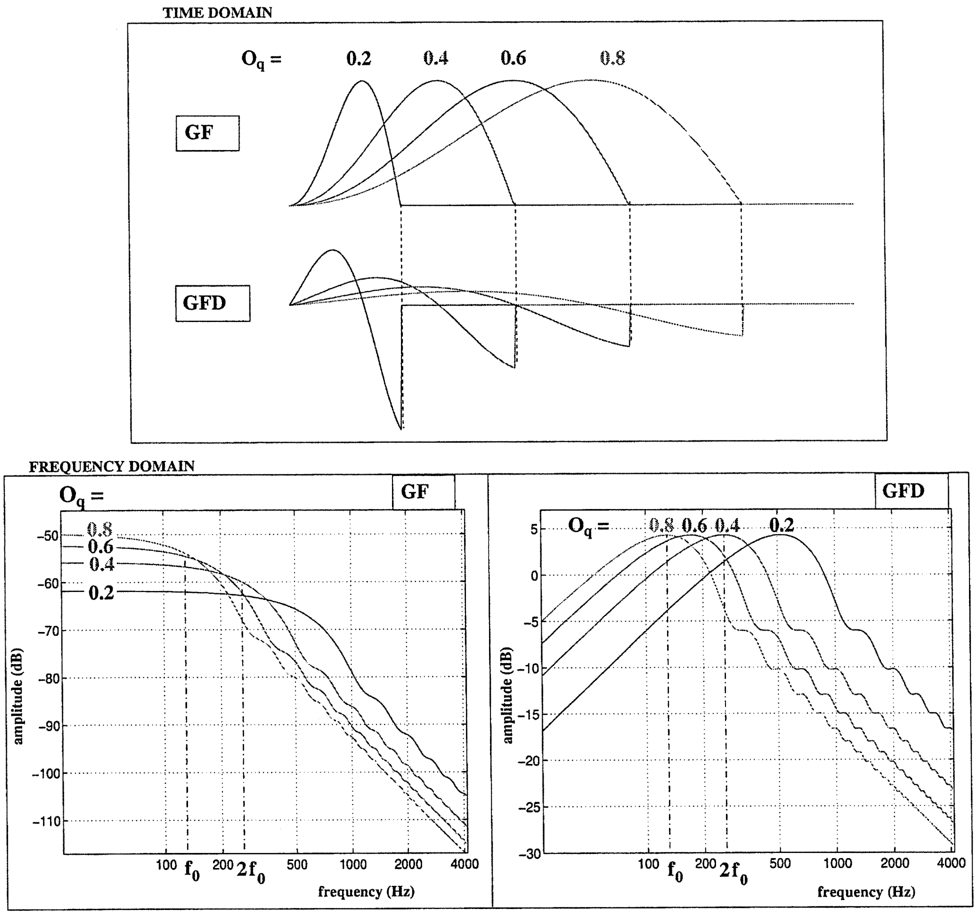

FIGURE 4. Open quotient variations shown in time and frequency domain. GF: glottal flow. GFD: glottal flow derivative. The frequency scale is logarithmic.

asked to choose which sound of the three was different. The threshold was decreased after two correct answers and increased after one incorrect answer. A series of increasing or decreasing threshold is defined as a run. During the whole test, no feedback was given to the subject with regard to the correct answer. The values of the initial threshold level, the initial step, and the last step were determined in a pilot study and are given in Table 1. Every second run, the initial step was divided by two, until it reached the last step value (0.01). The session was concluded after 14 runs in the case of open quotient and 12 runs in the case of asymmetry coefficient. The last eight runs were used for the calculation of the JND value for the given session.

The listening test took place in a sound isolation booth. The intensity level was adjusted so that the listeners would feel comfortable (approximately 70–80 dBA). The level was kept fixed for all subjects.

RESULTS

Learning effect

Ten of the 20 subjects who passed the screening repeated the listening test three times. A paired samples *t*-test[26] was conducted on their results to evaluate the learning effect from one test to another. When comparing the first and second tests, differences were found to be very significant for session 1 ($p < 0.01$) and significant ($p < 0.05$) for sessions 2, 4, 8, and 9. No statistical differences were found between sessions when comparing the second and third test. According to these results, we could assume that there is a learning effect between the

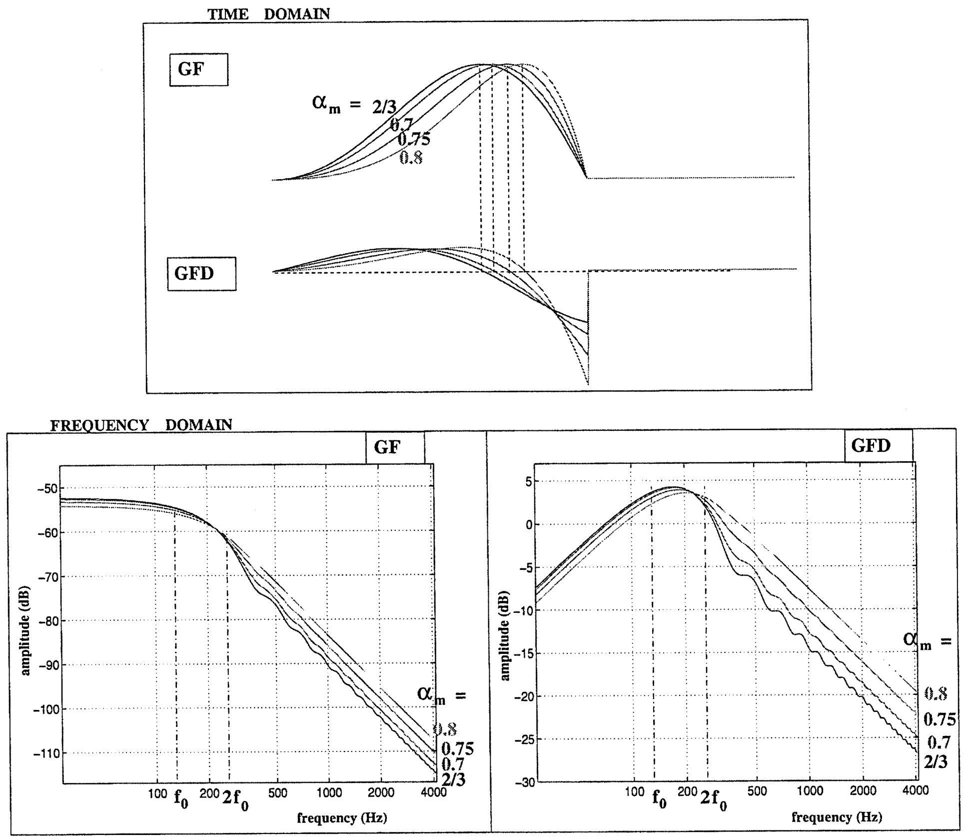

FIGURE 5. Asymmetry coefficient variations shown in time and frequency domain. GF: glottal flow. GFD: glottal flow derivative. The frequency scale is logarithmic.

first and second test and that the results obtained between the second and third test are comparable. These results are in agreement with those obtained by Scherer et al,[13] who observed that the first test yielded the highest JND values.

In order to take the learning effect into account in the data analysis process, subjects are divided into three groups as follows:

1. A group (Gl) of 20 subjects who are considered to be untrained. This group is composed of the

TABLE 1. *Details of the nine sessions. Precisions are given on the varied glottal flow parameter (either O_q or α_m), on the fundamental frequency and selected vowel. The initial step and threshold are given. The last step is 0.01 for each session. The default parameter values are $A_v = 1$, $O_q = 0.6$, $\alpha_m = 2/3$, $f_0 = 196$ Hz*

session	variable	O_q	α_m	f_0	vowel	vibrato	E	initial threshold	initial step
1	O_q	0.4	2/3	196 *Hz*	[a]			0.19	0.08
2	O_q	0.6	2/3	196 *Hz*	[a]	no vibrato		0.25	0.08
3	O_q	0.6	2/3	196 *Hz*	[a]			0.25	0.08
4	O_q	0.6	2/3	196 *Hz*	[a]		E fixed	0.25	0.08
5	O_q	0.6	2/3	130 *Hz*	[a]			0.25	0.08
6	O_q	0.8	2/3	196 *Hz*	[a]			0.19	0.08
7	α_m	0.6	2/3	196 *Hz*	[a]			0.13	0.04
8	α_m	0.6	0.8	196 *Hz*	[a]			0.13	0.04
9	O_q	0.6	2/3	196 *Hz*	[e]			0.25	0.08

10 subjects who took the test once and the 10 subjects who took the test three times but for whom only the first test is taken into account.

2. A group (G2) of 10 trained subjects. Only the 10 subjects who repeated the tests three times are considered. The considered threshold is calculated as the mean of the results of the second and third tests. The results of the first test are discarded.
3. A group (G3) of 7 subjects who did not pass the screening.

Mean and extreme thresholds

Table 2 gives the JND results for the three groups of subjects and the nine sessions. These results are in agreement with those obtained by Scherer et al.[13] In the case of the open quotient, they found a JND of 0.0344 for $O_q = 0.6$ and $f_o = 130.39$ Hz. These results are only slightly smaller than the JND of 0.041 obtained by the group G2 in session 2. In the case of the speed quotient, they found a JND of 0.319 for $S_q = 2$, which corresponds to a JND of 0.031 for $\alpha_m = 2/3$. This result is slightly higher than the JND of 0.027 obtained by the group G2 in session 7. However, it should be noticed that the pitch differs and that there is a vibrato in our case.

Noticeable differences were found among the three groups. Group G3 always had the highest JND for the mean value ($p < 0.01$) or the min value ($p < 0.01$). On the contrary, the trained subjects (group G2) had the best results for the mean JND ($p < 0.01$) and the max JND ($p < 0.01$). There was no significant difference on the min JND between untrained and trained subjects.

Open quotient

The results for the three values of open quotient $O_q = 0.4$, 0.6, and 0.8 are plotted in Figure 6 for group G1 and in Figure 7 for group G2.

The threshold is dependent on the target value. In the case of untrained subjects, the JND varies from 0.058 for $O_q = 0.4$, to 0.087 for $O_q = 0.6$ and 0.106 for $O_q = 0.8$. When the training is taken into account, it varies from 0.037 for $O_q = 0.4$, to 0.063 for $O_q = 0.6$, and to 0.079 for $O_q = 0.8$. The training effect can be seen in Figure 7, especially for subjects

TABLE 2. *Results for the three groups: mean value, standard deviation, minimum, and maximum value of thresholds (JND) for the nine sessions*

JND	Session 1 $O_q = 0.4$	3 $O_q = 0.6$	6 $O_q = 0.8$	2 no vib.	4 E const.	5 $f_0 = 130$	9 vowel [e]	7 $\alpha_m = 2/3$	8 $\alpha_m = 0.8$
GROUP G1 OF 20 UNTRAINED SUBJECTS									
mean	0.058	0.087	0.106	0.068	0.173	0.092	0.080	0.033	0.027
sd	0.019	0.027	0.039	0.032	0.077	0.036	0.029	0.009	0.009
min	0.022	0.043	0.049	0.009	0.034	0.049	0.042	0.019	0.012
max	0.092	0.153	0.190	0.161	0.298	0.190	0.144	0.052	0.049
GROUP G2 OF 10 TRAINED SUBJECTS									
mean	0.037	0.063	0.079	0.041	0.099	0.071	0.053	0.027	0.022
sd	0.010	0.015	0.014	0.008	0.047	0.018	0.013	0.010	0.008
min	0.024	0.043	0.061	0.027	0.044	0.051	0.033	0.013	0.012
max	0.053	0.090	0.101	0.052	0.178	0.093	0.067	0.048	0.037
GROUP G3 OF 7 SUJETS PRESENTING AUDITIVE WEAKNESS									
mean	0.063	0.133	0.156	0.075	0.235	0.149	0.142	0.041	0.032
sd	0.012	0.029	0.017	0.019	0.087	0.052	0.031	0.008	0.012
min	0.048	0.099	0.128	0.051	0.118	0.088	0.100	0.030	0.010
max	0.081	0.180	0.173	0.106	0.356	0.224	0.198	0.049	0.044

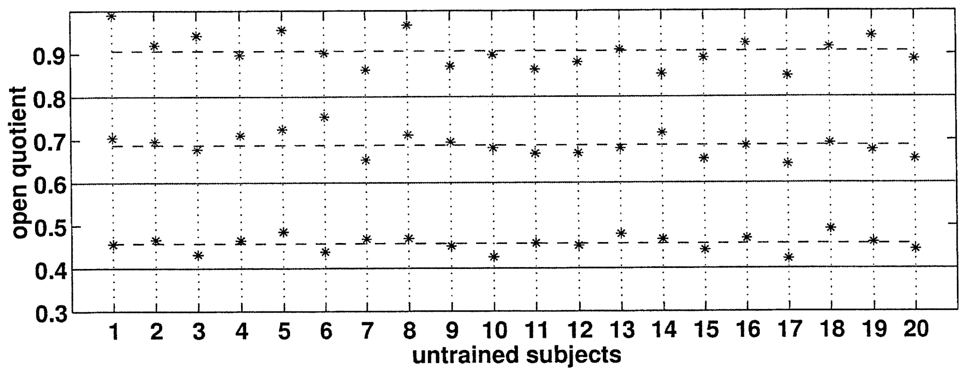

FIGURE 6. Open quotient values corresponding to the JNDs measured for three values of O_q (0.4, 0.6, and 0.8), in the case of 20 untrained subjects. The solid lines give the target values and the dashed lines the mean JND.

S16, S18, and S19, for whom the results of test 1 are much higher than those obtained in test 2 or 3. Table 3 shows the mean value of JND variations calculated between two sessions, showing comparisons among the three sessions (1, 3, and 6) where the open quotient was varied. It was found that session 1, where $O_q = 0.4$, was highly significantly different from the other two. The difference between session 3 ($O_q = 0.6$) and session 6 ($O_q = 0.8$) was less but still significant.

Asymmetry coefficient

The results for the two values of asymmetry coefficient, $\alpha_m = 2/3$ and 0.8, are plotted in Figure 8 for group G1 and in Figure 9 for group G2.

Contrary to the results obtained for open quotient, the JND seems to be less affected by a modification of the asymmetry coefficient target value. It goes from 0.033 for $\alpha_m = 2/3$ to 0.027 for $\alpha_m = 0.8$ in the case of group G1 and from 0.027 for $\alpha_m = 2/3$ to 0.022 for $\alpha_m = 0.8$ in the case of group G2. Table 3 shows that the difference between the two sessions is just significant.

Effects of vowel, pitch, vibrato, and amplitude

One question this study aims to answer is whether a given JND of open quotient would be dependent on the choice of vowel, pitch, amplitude parameter, and the presence of vibrato. Table 4 shows the comparison between the corresponding two sessions for each case for the open quotient.

No significant difference is found when the vowel is changed from [AA] to [IY] for group Gl, but a significant difference is found for group G2. This implies that with training, a slight effect of the vowel can be observed. In any case, the mean threshold is

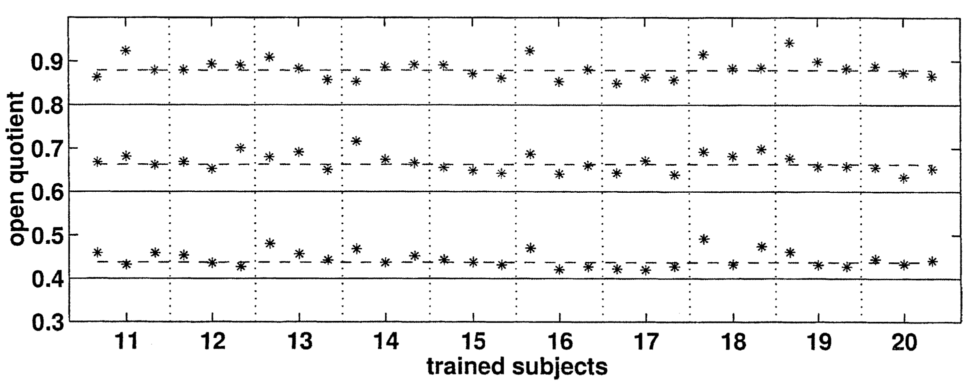

FIGURE 7. Open quotient values corresponding to the JNDs measured for three values of O_q (0.4, 0.6, and 0.8), in the case of 10 trained subjects (subjects S11 to S20). The solid lines give the target values and the dashed lines the mean JND.

TABLE 3. *Comparison between two given sessions in the case of O_q or α_m target value modification for untrained and trained subjects. For each group, the mean value of the threshold difference between two sessions is given (in %). A significant difference is marked by * ($p < 0.05$), a very significant difference by ** ($p < 0.01$), and a highly significant difference by *** ($p < 0.001$).*

comparisons	$O_q = 0.4 \leftrightarrow 0.6$ sessions $1 \leftrightarrow 3$	$O_q = 0.6 \leftrightarrow 0.8$ sessions $3 \leftrightarrow 6$	$O_q = 0.4 \leftrightarrow 0.8$ sessions $1 \leftrightarrow 6$	$\alpha_m = 2/3 \leftrightarrow 0.8$ sessions $7 \leftrightarrow 8$
untrained subjects	−2.96***	−1.86*	−4.81***	0.56*
trained subjects	−2.59***	−1.59**	−4.18***	0.58*

slightly smaller in the case of vowel [IY] than in the case of vowel [AA]. The pitch does not seem to have any effect on the JND value, and the results are therefore quite similar for $f_0 = 130$ Hz and 196 Hz.

On the other hand, a slight variation of frequency, corresponding to the natural vibrato, seems to have an effect on the JND values and its effect is even more significant for the trained ($p < 0.01$) than for the untrained ($p < 0.05$) subjects. The use of a vibrato helps to give some kind of naturalness to the synthetic sounds, and the JND is then increased.

Not surprisingly, the most important effect is due to the choice of amplitude parameter, as it directly affects the SPL. When E is kept fixed, the JND is significantly increased ($p < 0.001$) in the case of untrained subjects. The difference is less important but still significant for the trained subjects ($p < 0.05$).

DISCUSSION

The results have shown that the JND of open quotient is dependent on the value around which it is determined: the higher the open quotient, the higher the JND measure. A trend has also been found for the JND of asymmetry coefficient: the higher the asymmetry coefficient, the lower the JND. Moreover, the measure of the JND of open quotient is mainly affected by the choice of an amplitude parameter, as it greatly increases when the parameter E is kept constant. It is also dependent on the absence of a vibrato, as it decreases if there is no vibrato. No effect was found with regard to the chosen pitches, and only a slight effect was noticed for the trained group with regard to the chosen vowels. In other words, we should better perceive a given variation of open quotient: (1) around a low open quotient value, (2) if there is no vibrato, and (3) when the amplitude parameter that is kept constant is A_v, so that E is free to change with the open quotient. We should better perceive a given variation of asymmetry coefficient around a high asymmetry coefficient value.

JND of open quotient

The JND of open quotient is highly dependent on the target value and increases with an increase

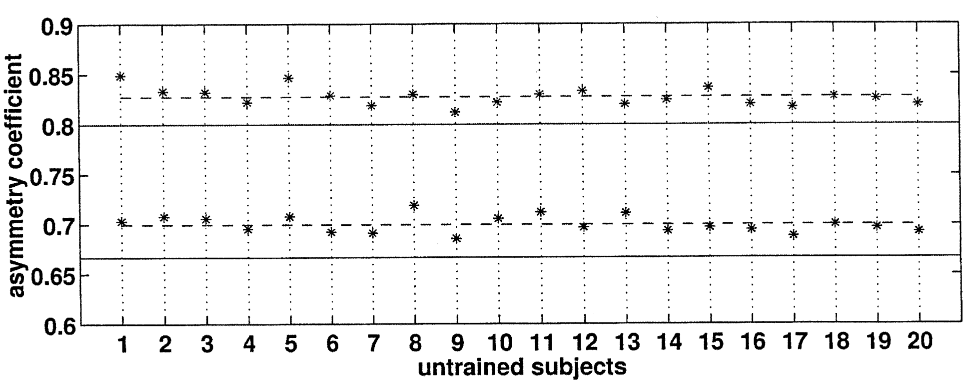

FIGURE 8. Asymmetry coefficient values corresponding to the JNDs measured for two values of α_m (2/3 and 0.8), in the case of 20 untrained subjects. The solid lines give the target values and the dashed lines the mean JND.

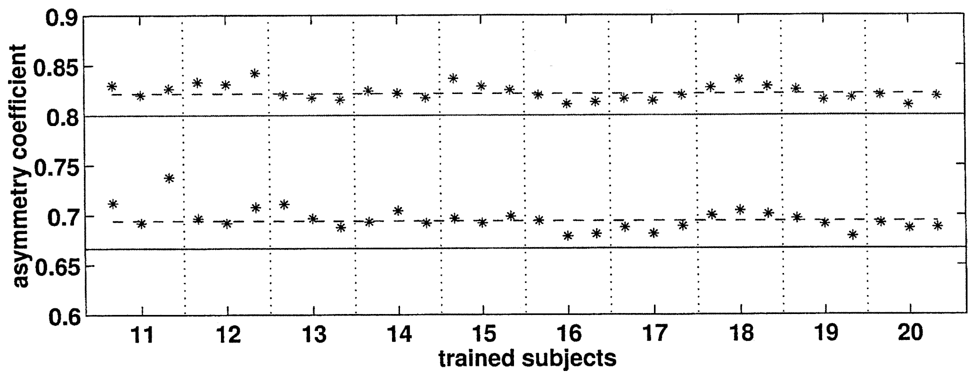

FIGURE 9. Asymmetry coefficient values corresponding to the JNDs measured for two values of α_m (2/3 and 0.8), in the case of 10 trained subjects (subjects S11 to S20). The solid lines give the target values and the dashed lines the mean JND.

of open quotient. What about the relative JND of open quotient, defined as the ratio $\Delta O_q/O_q$? Table 5 shows the relative JNDs for the three groups and the nine sessions, calculated as $\Delta O_q/O_q$ in the case of open quotient and $\Delta\alpha_m/\alpha_m$ in the case of asymmetry coefficient. The differences among session 1 ($O_q = 0.4$), session 3 (0.6), and session 6 (0.8) are not significant anymore with regard to the relative JND measures ($p > 0.05$ for each comparison). This experiment should be extended to more than only three open quotient values, but a trend is already shown:

1. In the case of the untrained group, $\Delta O_q \backsimeq 0.14 Oq$.
2. In the case of the trained group, $\Delta O_q \backsimeq 0.1\ Oq$.

How could we explain these results? We believe that this behavior is mainly due to frequency modifications of the glottal flow spectrum in the spectral region located within and above the first formant ($F_1 = 560$ Hz in our case), which are induced by the variations of the negative peak amplitude of the differentiated glottal flow (E). Indeed, the spectral slope is proportional to this parameter E.[14] When the amplitude of voicing (A_v) is kept constant, as well as the other glottal flow parameters, a variation of open quotient involves a variation of E according to Equation 4 and as previously illustrated in Figure 4. Thus, on a theoretical point of view, the relative JND of open quotient is then directly linked to the one of E by:

$$\Delta E/E = \Delta O_q/O_q$$

The parameter E is linked to the vocal intensity.[2–4,27] Thus, the JNDs of E should then be in agreement with the Bouguer–Weber law[28] that states that a JND (for example, ΔE in the case of vocal intensity variations) is proportional to the stimulus level (E) and, therefore, that the ratio $\Delta E/E$ is constant. Our results show that the relative JNDs of open quotient (and by extension of E) are actually in agreement with the Bouguer–Weber law. Thus, the JNDs we have

TABLE 4. *Comparison for the JNDs of open quotient between two given sessions in the case of a vowel, pitch, vibrato, or amplitude factor modification for untrained and trained subjects. For each group the mean value of the threshold difference between two sessions is given (in %). A significant difference is marked by * ($p < 0.05$), a very significant difference by ** ($p < 0.01$), and a highly significant difference by *** ($p < 0.001$).*

comparisons	vowel sessions 3 ↔ 9	pitch sessions 3 ↔ 5	vibrato sessions 3 ↔ 2	A_v / E sessions 3 ↔ 4
untrained subjects	0.75	−0.42	1.91*	−8.55***
trained subjects	1.01*	−0.80	2.18**	−3.57*

TABLE 5. *Relative JNDs for the 3 groups (G1: untrained subjects, G2: trained subjects and G3: subjects presenting auditive weakness) and the 9 sessions. (SD): standard deviation.*

Session	1 $O_q = 0.4$	3 $O_q = 0.6$	6 $O_q = 0.8$	2 no vib.	4 E const.	5 $f_0 = 130$	9 vowel [e]	7 $\alpha_m = 2/3$	8 $\alpha_m = 0.8$
G1	14.5	14.6	13.2	11.4	28.8	15.3	13.3	5.0	3.4
(SD)	(4.7)	(4.5)	(4.8)	(5.4)	(12.8)	(6.0)	(4.7)	(1.3)	(1.1)
G2	9.3	10.5	9.9	6.9	16.5	11.9	8.9	4.1	2.7
(SD)	(2.6)	(2.5)	(1.7)	(1.3)	(7.8)	(3.0)	(2.2)	(1.5)	(1.0)
G3	15.7	22.1	19.5	12.5	39.1	24.8	23.6	6.2	4.0
(SD)	(3.0)	(4.9)	(2.1)	(3.2)	(14.4)	(8.7)	(5.2)	(1.1)	(1.5)

measured in this study may mainly correspond to the perception of vocal intensity variations. In addition, the high values of JNDs obtained by the seven subjects presenting auditive weakness at high frequencies (ranging between 2000 Hz and 8000 Hz) further support that the JNDs of open quotient are dependent on high-frequency spectral variations.

JND of asymmetry coefficient

As with the case of the open quotient, the parameter E varies with the asymmetry coefficient, when all other parameters are kept fixed. The relation between both parameters is no longer simple, because of the model implicit equation. On Figure 10, the parameter E is plotted as a function of α_m. This curve

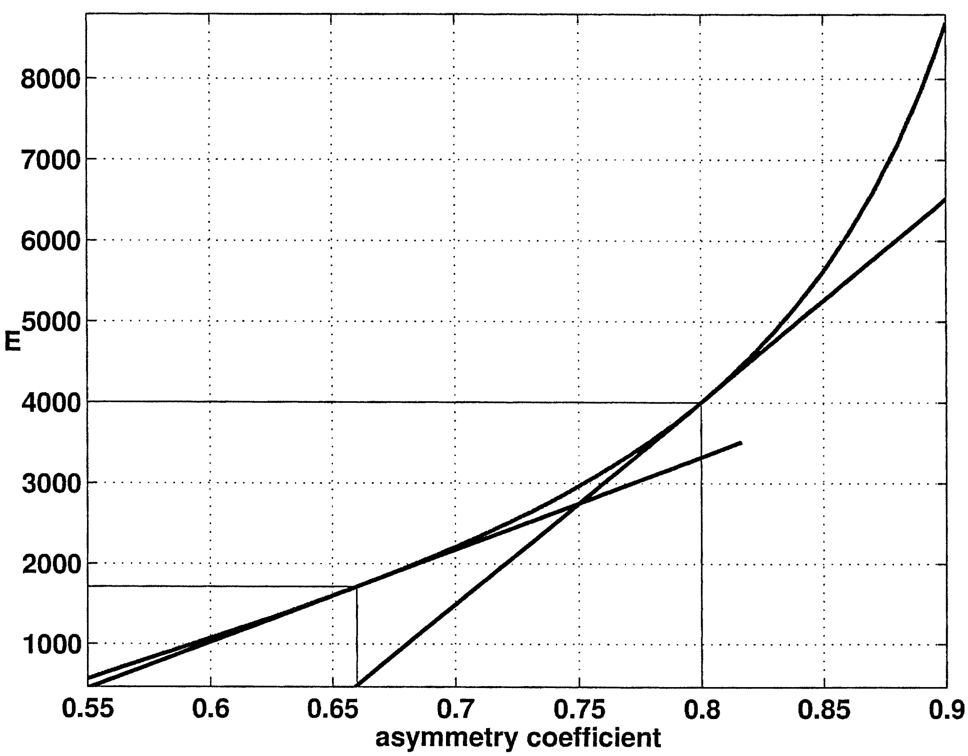

FIGURE 10. Variations of the negative peak amplitude of the differentiated glottal flow E as a function of asymmetry coefficient. The dafault values are those of session 3: $A_v = 1$, $f_0 = 196$ Hz, $O_q = 0.6$, and $\alpha_m = 2/3$. The curve's tangents are plotted for $\alpha_m = 2/3$ and 0.8.

corresponds to numerical solutions of Equation 4 for α_m ranging between 0.55 and 0.9. The curve's tangent is plotted for $\alpha_m = 2/3$ and 0.8. The tangent slope increases when α_m increases, which means that the variation of E (and therefore of vocal intensity) related to a given variation of α_m will be greater for high values of the asymmetry coefficient. This may explain the behavior we observed, a decrease of JND as α_m increases.

The relative JNDs obtained for the asymmetry coefficient are shown in Table 5 for the two sessions and for the three groups. A difference was observed between the two sessions, which was found to be very significant for each group ($p < 0.01$). Thus, contrary to the open quotient, the relative JND of asymmetry coefficient is not constant over the two sessions and it decreases when α_m increases. Table 6 gives the corresponding values of $\Delta E/E$, which cannot be expressed simply as a function of $\Delta\alpha_m/\alpha_m$ and were thus calculated numerically. For both sessions, greater values are found when compared to the results of the open quotient. This difference may be explained by differences in low-frequency spectral variations. As we already mentioned, the glottal formant frequency is shifted when O_q is varied. A variation of α_m does not affect strongly this glottal formant frequency but mainly its bandwith and therefore the low-frequency variations will be smaller and less perceived in the case of the asymmetry coefficient than in the case of the open quotient.

CONCLUSION

The variations of glottal flow parameters affect voice quality. Yet, only a few studies were conducted to quantify the relationship between vocal production and vocal perception. The purpose of this study is to measure to which extent a given glottal flow parameter could be varied without being perceived. It is a sequel to a study done by Scherer et al,[13] who measured Just Noticeable Differences of open quotient and speed quotient around one target value of these glottal flow parameters, at a fixed fundamental frequency and for a given vowel. We measured the JNDs for three values of open quotient and two values of asymmetry coefficient (equivalent to the speed quotient parameter), and we explored the effects on the JND of open quotient of the presence of a vibrato, of the chosen vowel, fundamental frequency, and amplitude parameter.

The results obtained in a comparative session showed good agreements with the results obtained by Scherer et al.[13] It was found that the JND of open quotient is dependent on the target value and increases when open quotient increases. When considering the relative JND, it was found to be relatively stable within a group, being approximately 14% for the untrained group and 10% for the trained group. These results indicate also that only great variations of open quotient will be perceived. In the case of asymmetry coefficient, the JND is also dependent on the target value, as it decreases when asymmetry coefficient increases. In this case, the relative JND is no longer constant. But it should be noted that only two values of asymmetry coefficient have been explored in this study. The relative JND of asymmetry coefficient is of the order of 4%, which means that a small variation of this parameter will be easily perceived. Comparing two different sessions for each case, no effect was found on the JND of open quotient when choosing vowel or fundamental frequency. The JND increases in the presence of a natural vibrato or when the negative peak amplitude of the differentiated glottal flow is kept fixed.

The perception of variations of open quotient or asymmetry coefficient seems to depend mainly on glottal spectral slope amplitude modifications that result from the underlying variations of the negative peak amplitude of the differentiated glottal flow. It would be of interest to extend the present work to measure the JNDs while keeping the parameter E fixed, not only in one session as done here, but

TABLE 6. *Relative JNDs for E in the case of a variation of asymmetry coefficient (sessions 7 and 8) and for the 3 groups (G1: untrained subjects, G2: trained subjects and G3: subjects presenting auditive weakness). (SD): standard deviation*

Session	7 $\alpha_m = 2/3$	8 $\alpha_m = 0.8$
G1	23.1	19.9
(SD)	(6.7)	(7.9)
G2	19.0	15.2
(SD)	(7.7)	(6.3)
G3	29.2	23.8
(SD)	(5.8)	(9.8)

throughout the entire test, so as to characterize the low-frequency perception of open quotient and asymmetry coefficient apart from vocal intensity variations due to spectral slope amplitude modifications.

These results provide useful information for speech and singing synthesis. As an example, the variation step of a given parameter could be chosen with regard to the relative JND measures. In the case of voice source analysis, two different physiological measures could be compared along a JND-based perceptual axis. Continuing research will then provide the possibility of relating directly the production of a sound to its perception, which would be very useful for voice analysis and synthesis, vocal and ear training, as well as for clinical purposes.

REFERENCES

1. Monsen RB, Engebretson AM. Study of variations in the male and female glottal wave. *J Acous Soc Am*. 1977;62: 981–993.
2. Holmberg EB, Hillman RE, Perkell JS. Glottal air flow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice. *J Acous Soc Am*. 1988;84:511–529.
3. Gauffin J, Sundberg J. Spectral correlates of glottal voice source waveform characteristics. *J Speech Hear Res*. 1989; 32:556–565.
4. Sundberg J, Titze IR, Scherer RC. 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*. 1993;7:15–29.
5. Timcke R, yon Leden H, Moore P. Laryngeal vibrations: measurements of the glottic wave. *AMA Arch Otolaryngol*. 1958;68:1–19.
6. Orlikoff RF. Assessment of the dynamics of vocal fold contact from the electroglottogram: data from normal male subjects. *J Speech Hear Res*. 1991;34:1066–1072.
7. Dromey C, Stathopoulos ET, Sapienza CM. Glottal airflow and electroglottographic measures of vocal function at multiple intensities. *J Voice*. 1992;6(1):44–54.
8. Hacki T. Electroglottographic quasi-open quotient and amplitude in crescendo phonation. *J Voice*. 1996;10:342–347.
9. Moore BCJ, Glasberg BR, Baer T. A model for the prediction of thresholds, loudness and partial loudness. *J Audio Eng Soc*. 1997;45:224–240.
10. Veldhuis R. The spectral relevance of glottal-pulse parameters. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*. Seattle, USA; 1998:873-876.
11. Rao P, van Dinther R, Veldhuis R, Kohlrausch A. A measure for predicting audibility discrimination thresholds for spectral envelope distortions in vowel sounds. *J Acous Soc Am*. 2001;109:2085–2097.
12. van Dinther R, Veldhuis R, and Kohlrausch A. The perceptual relevance of glottal-pulse parameter variations. In *Proc. Eurospeech 2001*, Aalborg, Denmark, Sept. 2001.
13. Scherer RC, Hoberg Arehart K, Gent Guo C, Milstein CF, Horii Y. Just Noticeable Differences for glottal flow waveform characteristics. *J Voice*. 1998;12(1):21–30.
14. Doval B, d'Alessandro C. The spectrum of glottal flow models. *notes et documents LIMSI*, 99-07, 1999.
15. Titze IR, Mapes S, Story B. Acoustics of the tenor high voice. *J Acous Soc Am*. 1994;95(2):1133–1142.
16. Vu Ngoc T, d'Alessandro C. Robust glottal closure detection using the wavelet transform. In *Proc. Eurospeech 1999*, pp. 2805-2808, Budapest, Hungary, Sept. 1999.
17. Vu Ngoc T, al'Alessandro C. Glottal closure detection using EGG and the wavelet transform. In *Proc. 4th Intern. Workshop on Advances in Quantitative Laryngoscopy, Voice and Speech Research*, Jena, Germany, Apr. 2000.
18. Henrich N. Etude de la source glottique en voix parlé et chanté. (Study of the glottal source in speech and singing). Doctoral thesis. France: Universit Paris 6; 2001.
19. Fant G, Liljencrants J, Lin Q. A four-parameter model of glottal flow. *STL-QPSR*. 1985;4:1–13.
20. Shoup JE. Phonological aspects of speech recognition. In *Trends in Speech Recognition*. Englewood Cliffs, NJ: Prentice Hall; 1980:125–138.
21. Doval B. and d'Alessandro C. Spectral correlates of glottal waveform models: An analytic study. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*. Munich, Germany, April 1997.
22. Henrich N, d'Alessandro C, Doval B. Spectral correlates of voice open quotient and glottal flow asymmetry: theory, limits, and experimental data. In *Eurospeech 2001*, Aalborg, Denmark, September 2001.
23. Fant G. Glottal source and excitation analysis. *STL-QPSR*. 1979;1:85–107.
24. Levitt H. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*. 1971;49(2):467–477.
25. Rioux V. Sound quality of flue organ pipes—An interdisciplinary study of the art of voicing. Doctoral Thesis. Gothenburg, Sweden: Chalmers University of Technology; 2001.
26. Daudin JJ, Robin S, Vuillet C. Statistique infrentielle. Ides, dmarches, exemples. *Pratique de la statistique—Socit Franaise de Statistique et Presses Universitaires de Rennes*. 1999.
27. Fant G. Preliminaries to analysis of the human voice source. *STL-QPSR*. 1982;4:1–27.
28. Bonnet C. *Manuel pratique de psychophysique*. Paris, France: Armand Colin; 1986.