



A method for analysing the perceptual relevance of glottal-pulse parameter variations

R. van Dinther^{a,*}, A. Kohlrausch^{a,b}, R. Veldhuis^c

^a Department of Technology Management, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

^b Philips Research Laboratories Eindhoven, Prof. Holstlaan 4, WY 82, 5656 AA Eindhoven, The Netherlands

^c Faculty of Engineering, Chair of Signals and Systems, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Received 17 June 2002; received in revised form 5 July 2003; accepted 29 July 2003

Abstract

This paper describes a method for analysing the perceptual relevance of parameter variations of the Liljencrants–Fant (LF) model for the glottal pulse. A perceptual distance measure based on excitation patterns was developed and evaluated in order to predict audibility discrimination thresholds for small changes to the R -parameters of the LF model. For a number of R -parameter sets, taken from real source data, an approximation Q of the distance measure was used to compute the directions of maximal and minimal perceptual sensitivity. In addition, we show that the inverse of Q can be used to calculate the amount of variation of the R -parameters to reach a just noticeable difference (JND). The results were evaluated in a listening test in which JNDs for R -parameter changes were measured. Discrimination thresholds were fairly constant across all tested conditions and corresponded on average to an excitation pattern distance of 4.3 dB. An additional error analysis demonstrated that Q is a fair approximation of the perceptual distance measure for small variations of the R -parameters up to one just noticeable difference.

© 2003 Elsevier B.V. All rights reserved.

Zusammenfassung

In dieser Arbeit wird eine Methode beschrieben, die es erlaubt, die perzeptive Relevanz von Parametervariationen des Liljencrants–Fant (LF) Modells für den glottalen Puls zu bestimmen. Hierzu wurde ein perzeptives Abstandsmaß entwickelt, das auf Erregungspegelmustern aufbaut. Die Qualität dieses Maßes wurde mit Hilfe von experimentellen Daten zur Wahrnehmbarkeit kleiner Veränderungen in den R Parametern des LF Modells getestet. Für eine Anzahl von R -Parameter Vektoren, die aus Sprachäußerungen berechnet waren, wurde eine Approximation Q des Abstandsmaßes verwendet, um die Richtungen größter sowie kleinster perzeptiver Empfindlichkeit zu bestimmen. Darüberhinaus wird gezeigt, dass die Inverse von Q benutzt werden kann, um die Veränderung in den R Parametern berechnen zu können, die zu einer gerade wahrnehmbaren Änderung (JND) führt. Diese Ergebnisse wurden anhand von Experimenten überprüft, in denen direkt die JND Werte für Veränderungen in den verschiedenen R Parametern gemessen wurden. Die Unterschiedsschwellen waren in allen getesteten Situationen relativ konstant und der Mittelwert entsprach einem

* Corresponding author. Address: Centre for the Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK.

E-mail address: ralph.van-dinther@mrc-cbu.cam.ac.uk (R. van Dinther).

Unterschied im Erregungsmuster von 4.3 dB. Eine Fehleranalyse zeigte, dass Q eine gute Näherung des perceptiven Abstandsmaßes darstellt, zumindest für kleine Unterschiede bis zur Größe des gerade wahrnehmbaren Unterschiedes.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Speech synthesis; Glottal-pulse; LF parameters; Perceptual distance measure

1. Introduction

Many studies in speech synthesis are concerned with the parameterisation of the voice source. A considerable amount of work has been done to derive glottal-flow characteristics by methods such as inverse filtering and analysis in the frequency domain. These methods have provided insight into human speech production and various ways to quantify the glottal-flow by suitable parametric systems. In voiced speech the glottal-flow is (quasi) periodic. One period is called the glottal pulse. One of the most widely used models to quantify the glottal-pulse is the Liljencrants–Fant (LF) model (Fant et al., 1985). In this model the shape of the glottal-pulse is characterised by three parameters, here called the *R*-parameters. The *R*-parameters, together with the fundamental frequency and an amplitude parameter, completely determine the glottal-pulse.

The relation between glottal-pulse parameters and voice quality has been investigated in a number of studies (Gobl, 1989; Karlsson, 1990; Childers and Lee, 1991; Gobl and Ní Chasaide, 1992; Karlsson and Liljencrants, 1996). However, little attention has so far been paid to the perceptual relevance of glottal-pulse parameter variations. Experimental assessment of the perceptual relevance of glottal-pulse parameters would require a very extensive experiment: one would have to include a number of vowel conditions, a number of glottal-pulse parameter sets and one would have to vary these parameters systematically along various directions in the parameter space. An approach like this has been followed by Doval and d'Alessandro (1997), Doval et al. (1997), Scherer et al. (1998) and Henrich et al. (in press).

Such extensive experimental research can be circumvented by combining two results from earlier studies. First, Veldhuis (1998) investigated the

spectral relevance of the *R*-parameters of the LF model. His study showed that the spectrum is hardly affected by certain variations of the *R*-parameters. Second, in (Rao et al., 2001) it was shown that the representation of an auditory stimulus in a perceptual domain allows one to predict discrimination thresholds between similar acoustic stimuli. In particular, the difference between excitation patterns (excitation pattern distance, EPD) or, alternatively, the partial loudness measure which is derived from a more advanced stage of the loudness model of Moore et al. (1997), can be used for this purpose. Excitation patterns approximate the representations of stationary sounds in the inner ear. Applying the analysis proposed by Veldhuis to a representation given in (Rao et al., 2001) opens up the possibility of computing the perceptual relevance of glottal-pulse parameter changes based on a validated perceptual model. In this way, extensive experiments can be avoided, although some calibration experiments may still be needed.

Following this line of thought, we present a new method of analysis which can be used to quantify the perceptual relevance of the parameters of the LF model at points in the parameter space. Although the analysis is general and can be presented for the parameters of various glottal-pulse models, we restrict ourselves to the *R*-parameters of the LF model.

We describe a mapping which allows us to compute the excitation pattern for each *R*-parameter set. Here the *R*-parameter sets are interpreted as points in a three-dimensional space. The perceptual distance between two *R*-parameter sets is based on the difference between the two excitation patterns corresponding to the two parameter sets. A local second-order approximation of this perceptual distance measure is used to compute the directions of maximum and minimum sensitivity

to R -parameter changes in this three-dimensional space for a number of R -parameter sets taken from real glottal-source data. The results are evaluated in a listening test in which the just noticeable differences (JNDs) for the R -parameter changes are determined.

Earlier work on the determination of JNDs for glottal-pulse parameter changes for both synthesised signals of the glottal-flow alone and vowel waveforms was done by Scherer et al. (1998) and Henrich et al. (in press). In these two studies JNDs were obtained for the parameters open quotient (O_q), speed quotient (S_q) and an asymmetry coefficient (α_m), which are closely related to the R -parameters. In the discussion of the experiments of the present study, we will compare the JNDs obtained in these studies to our JNDs and express them in terms of our EPD.

The structure of this article is as follows. Section 2 briefly discusses the LF model. In the following section the developed perceptual distance measure is described. The perceptual distance measure is examined by measuring the perceptual sensitivity in the experiments presented in Section 4 to changes of the three R -parameters. In Section 5 an approximation of D is introduced for the calculation of directions of maximum, intermediate and minimum perceptual sensitivity. A second listening experiment (Section 6) is used to evaluate D for variations to the R -parameters in these directions. In Section 7, the results of the experiments are discussed and are used to obtain estimates of the size of one JND. Section 8 presents conclusions.

2. The Liljencrants–Fant model

The LF model has become a reference model for glottal-pulse analysis. Fig. 1 shows as example one period of the glottal-pulse time derivative $g'(t)$, which is commonly used to model the source signal in a source-filter model of speech (Fant et al., 1985; Klatt and Klatt, 1990).

The duration of a glottal cycle is $T_0 = 1/F_0$, with F_0 the fundamental frequency. The maximum air-flow, often denoted as U_0 of the glottal-pulse occurs at T_p and the maximum excitation with

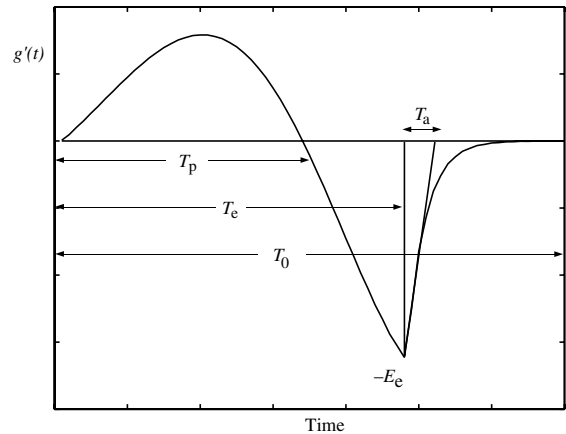


Fig. 1. Time derivative of one period of the glottal-flow and definition of glottal-pulse parameters.

amplitude E_c occurs at T_e , which corresponds to the instant of collision of the vocal cords. The interval before T_e is called the *open phase*. The interval with approximate length $T_a = E_c/g''(T_e)$ just after the instant of maximum excitation is called the *return phase*. The interval between $T_e + T_a$ and the end of the glottal cycle is called the *closed phase*. During this phase the vocal folds reach maximum closure and the airflow is reduced to its minimum. The minimum airflow is often referred to as leakage. Here we assume that there is no leakage, therefore $g(0) = g(T_0) = 0$. The airflow in the return phase is generally considered to be of perceptual importance, because it determines the spectral slope of the corresponding vowel. The parameters T_0 , T_p , T_e and T_a are called the T -parameters. In this paper we will use the related set of R -parameters, which are defined as $R_o = T_e/T_0$, $R_k = (T_e - T_p)/T_p$ and $R_a = T_a/T_0$. The parameters R_o and R_a denote the relative duration of the open phase and the return phase, respectively. The parameter R_k quantifies the symmetry of the glottal-pulse in the open phase. The mathematical relation between these R -parameters, the fundamental frequency, maximum airflow U_0 and maximum excitation E_c is given in (Fant et al., 1985). The *shape* of the glottal-pulse according to the LF model is fully specified by the parameters R_a , R_k and R_o . In this paper we will only study the effects of these shape parameters, which will be presented as

vectors in the following way: $\mathbf{r} := [R_a, R_k, R_o]^T \in \mathbb{R}^3$. The set of points \mathbf{r} will be denoted as the R -parameter space.

3. Mapping from the R -parameters into a perceptual space

In order to quantify R -parameter variations in a perceptual space, distance measures derived from a model for the auditory system are used. In (Rao et al., 2001), perceptual distance measures were developed and tested for changes in harmonic spectra. The measures are based on an auditory model presented in (Moore et al., 1997). It was shown in (Rao et al., 2001) that the partial loudness measure and the EPD are equally appropriate for predicting audibility discrimination thresholds. A mathematical difference between these two measures consists in the observation that the EPD is a smoother function of the harmonic amplitude spectrum than the partial loudness measure. Therefore, it can be better approximated by a quadratic functional and this is the reason why, in this paper, we use the EPD rather than partial loudness.

3.1. The perceptual distance measure D

The EPD described in (Rao et al., 2001) is based on a 5-stage model for computing loudness (Moore et al., 1997) which we adapted for representing amplitude changes in harmonic spectra. For the calculation of excitation patterns, only the first three stages of the loudness model are needed. The first two stages model transfer functions from the free field to the eardrum and through the middle ear. In the third stage, the excitation pattern of a given sound is calculated from the effective spectrum reaching the cochlea. According to Moore (1986), excitation patterns can be thought of as the distribution of “excitation” evoked by a particular sound in the inner ear along a frequency axis. In terms of a filter analogy, the excitation pattern represents the output levels of successive auditory filters as a function of their centre frequencies. Each auditory filter represents the frequency selectivity of the inner ear at a particular

frequency, where the filter shape varies with the input level. The excitation pattern is generally presented as a function of the ERB-rate rather than as a function of frequency. ERB stands for Equivalent Rectangular Bandwidth and is a measure for the bandwidth of the inner ear filters. The ERB scale is conceptually related to the Bark scale (Zwicker and Scharf, 1965; Moore, 1986). The ERB-rate is a value on the ERB scale, on which the auditory filters are uniformly spaced. The excitation patterns are continuous functions on an interval in \mathbb{R} . In this paper, the excitation patterns are calculated for ERB-rate values between 0 and 40. The ERB-rate has the following relation with the frequency scale:

$$\text{number of ERBs} := 9.26 \log(4.37F + 1), \quad (1)$$

where F is the frequency in kHz. Thus the range of 0–40 on the ERB scale corresponds to a frequency range of about 0–15 kHz. Because we want to measure the difference between two excitation patterns which are continuous functions on the interval (0, 40), we consider the excitation patterns as elements of the infinite-dimensional vector space $L^2(0, 40)$.¹ The excitation pattern distance (EPD) between two excitation patterns e_1 and e_2 , where the patterns are represented in dB, is quantified by

$$\|e_1 - e_2\|_2 := \left(\int_0^{40} |e_1(x) - e_2(x)|^2 dx \right)^{1/2}. \quad (2)$$

In this work, the integral is numerically approximated by a sum with $|e_1 - e_2|^2$ sampled at steps of 0.1 ERB.^{2,3}

The perceptual distance between two sounds will be denoted as the quantity dB EPD. The EPD will be used to determine the perceptual effect of small changes to the R -parameters. For a point \mathbf{r} in

¹ $L^2(0, 40)$ is the Banach space of Lebesgue integrable functions on the interval (0, 40).

² In (Rao et al., 2001) this numerical approximation of the distance between two excitation patterns was defined as the “Euclidean distance between the excitation patterns”, which suggests that the distance was taken between two vectors, where it was actually the distance between two densities.

³ The software in Matlab for the computation of excitation patterns can be obtained from the first author.

the R -parameter space the excitation pattern can be calculated from the power spectrum resulting from a combination of a glottal-pulse derivative according to the LF model and a filter modelling vowel formats. If necessary, the particular vowel used for the calculated excitation pattern will be indicated. We will denote the excitation pattern for a parameter vector \mathbf{r} as $e(\mathbf{r})$. The perceptual difference between the two parameter sets $\mathbf{r} + \mathbf{h}$ and \mathbf{r} is then quantified by the distance

$$D_r(\mathbf{h}) := \|e(\mathbf{r} + \mathbf{h}) - e(\mathbf{r})\|_2 \quad (3)$$

with $\mathbf{h} \in \mathbb{R}^3$. For convenience, whenever we refer to the distance measure in general, we will drop the index \mathbf{r} of D_r .

4. Experiment I: Evaluation of the distance measure D

This section describes Experiment I which was performed to evaluate whether D , as defined in Section 3, can be used to estimate audibility discrimination thresholds for small changes to the R -parameters. If so, the measure D can be used for our goal: the quantification of the perceptual relevance of the R -parameters.

In Experiment I the R_a , R_k and R_o are modified individually and, therefore, in orthogonal directions aligned with the primary axes of the R -parameter space. In the second experiment which will be presented in Section 6, the R -parameters are varied in other directions selected for their perceptual relevance. The waveforms of the glottal-pulse derivative corresponding to the five different R -parameter vectors used in the two experiments are shown in Fig. 2. Above each panel voice qualities corresponding to the glottal-pulses are given. The time axis covers a little more than one period of the glottal pulse.

4.1. Experiment I: Stimuli

The stationary vowels /a/ and /i/ were synthesised with a source-filter model. The number of harmonics was $N = 36$. The fundamental frequency of the vowels was $F_0 = 110$ Hz and the sampling frequency 8 kHz. Two parameter sets

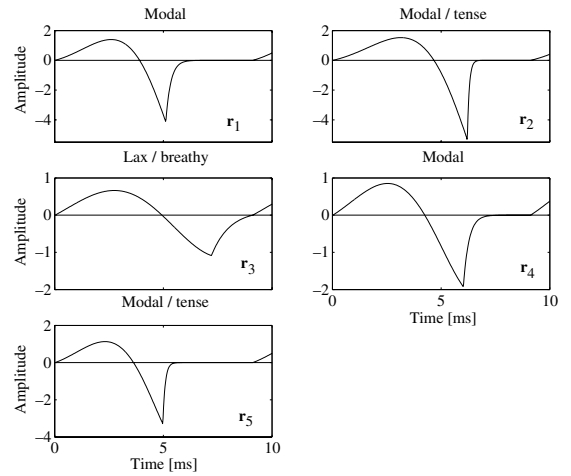


Fig. 2. Waveforms of the glottal-pulse time derivative for the 5 points $\mathbf{r}_1 := [0.025, 0.310, 0.560]^T$, $\mathbf{r}_2 := [0.010, 0.330, 0.680]^T$, $\mathbf{r}_3 := [0.092, 0.463, 0.791]^T$, $\mathbf{r}_4 := [0.028, 0.420, 0.660]^T$ and $\mathbf{r}_5 := [0.012, 0.374, 0.545]^T$ in the R -parameter space used in the experiments. Each waveform is characterised by its corresponding voice quality.

$\mathbf{r}_1 := [R_a, R_k, R_o]^T = [0.025, 0.310, 0.560]^T$ and $\mathbf{r}_2 := [0.010, 0.330, 0.680]^T$ were varied in three directions, corresponding to the principal axes of the R -parameter space, i.e. in the R_a , R_k and R_o directions. The flow amplitude U_0 , the maximum excitation E_e and the overall gain were allowed to vary with the R parameters. In Table 1, a scheme of the parameter conditions is presented. The symbols +/- indicate whether the specific R -parameter was increased (+) or decreased (-).

It is generally accepted that amplitude changes in the spectrum of harmonic sounds are more

Table 1
Parameter conditions of Experiment I showing the directions of modifications to the R -parameters $\mathbf{r}_1 := [0.025, 0.310, 0.560]^T$ and $\mathbf{r}_2 := [0.010, 0.330, 0.680]^T$, and the two different vowels

Direction	Conditions and corresponding modifications		
	/a/, \mathbf{r}_1	/a/, \mathbf{r}_2	/i/, \mathbf{r}_1
R_o	- (conditions 3, 4)		+ (conditions 11, 12)
R_a	+ (conditions 1, 2)	- (conditions 5, 6)	
R_k		+ (conditions 7, 8)	- (conditions 9, 10)

detectable than phase changes. For complex tones with a fundamental frequency beyond 150 Hz, the maximal effect of phase on timbre is smaller than the effect of changing the slope of the amplitude pattern by 2 dB/oct (Plomp and Steeneken, 1969). In the experiment we tested whether phase changes could be ignored for complex tones with a fundamental frequency of 110 Hz. For that purpose, two signal conditions were introduced. In one signal condition the phase was regular, i.e. the phase was allowed to change with the R -parameters. In the other signal condition the phase was fixed, i.e. after modifying the signal the phases were adjusted to those of the reference signal. In total there were 2 (signal conditions) \times 6 (parameter conditions) = 12 conditions. In the odd-numbered conditions the signals had regular phase.

4.2. Experiment I: Method

For each condition we computed the EPD between the excitation pattern of the reference sound and the excitation pattern of the modified sound for which the subjects were just able to discriminate between the two sounds. Four subjects (S1, S2, S3 and S4) participated in the experiments. All were young adults with normal hearing and no reported history of hearing impairment. The stimuli were presented diotically over headphones at a level of approximately 55 dB SPL. A 3-interval forced-choice adaptive procedure (Levitt, 1971) was used to obtain the thresholds. In this procedure each trial consisted of three stimuli, two stimuli representing the reference sound and one the modified sound. The assignment of the odd stimulus to one of the three intervals was randomised. The subject's task was to indicate the odd interval. Immediately after each response, feedback was given indicating whether the response was correct or incorrect. The interval after the response of the subject and the next trial was 300 ms. The inter-stimulus interval was 400 ms. After two correct responses the amount of modification of the R -parameters was reduced by one step. After one incorrect response it was increased by one step.

The R -parameter modifications of a stimulus were divided into 37 steps reaching from about 25

to 1.4 dB EPD for the modified sound. A run began with a modification of the R -parameters that produced a clearly discriminable change. A test run was completed after 12 up–down reversals. A single-run estimate of the EPD at threshold was obtained by taking the median of the EPDs corresponding to the steps within the last eight up–down reversals of the run. In this way the 71% correct detection threshold is measured. For each experimental condition the mean of four single-run estimates was taken as the final estimate of the distance for each subject.

4.3. Experiment I: Results

In Fig. 3 the actual JNDs in the R -parameter space of the subjects S1, S2, S3 and S4 are shown. On the y -axis a logarithmic scale is used. The diamonds and plus signs in the figures are the predicted JND values, calculated by the distance measures D and Q , which will be discussed in Section 7.

One can observe a division of the data into three groups according to the order of magnitude. These groups comprise the conditions 1, 2, 5 and 6; 7, 8, 9 and 10; and 3, 4, 11 and 12. The three dif-

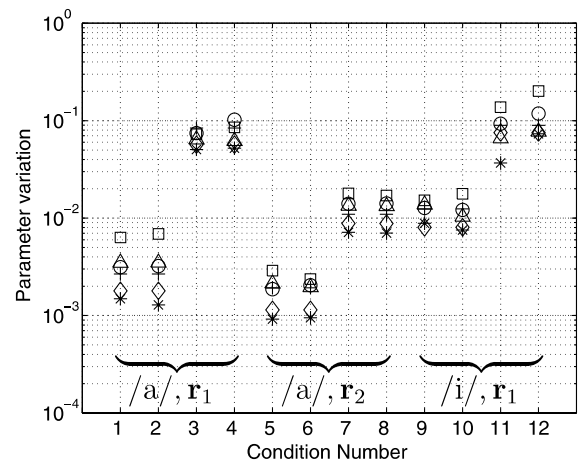


Fig. 3. Data obtained from Experiment I of the measured JNDs in the R -parameter space for each condition for points $r_1 := [0.025, 0.310, 0.560]^T$, $r_2 := [0.010, 0.330, 0.680]^T$. The diamonds indicate the predicted JND values $\tilde{\tau}_{i,\text{JND}}$ calculated with the inverse of D . The + signs indicate the predicted JND values $\tilde{\tau}_{i,\text{JND}}$ calculated with the inverse of Q . On the y -axes a logarithmic scale is used.

ferent orders of magnitude as listed above correspond to variations along the three parameter axes R_a , R_k and R_o , respectively. Comparing the conditions with regular and fixed phases (odd- and even-numbered conditions), it can be stated that there are no systematic phase effects. The figure also shows that variations in positive (e.g. conditions 1, 2) and negative directions (e.g. conditions 5, 6) give the same JND values. Finally, one can observe by comparing conditions 7, 8, 9 and 10 that the differences between the vowels /a/ and /i/ are negligible.

Fig. 4 shows the same data expressed in terms of the perceptual distance measure. The data of S1, S2, S3 and S4 are indicated with a square, a circle, an upward triangle and an asterisk, respectively. Ideally the curves in the figure would be flat if the measure was a perfect predictor and if subjects would have no variability in their responses.

The data presented in Fig. 4 show that the thresholds of the subjects lie between 1.5 and 10 dB EPD. The grand mean of the thresholds is 4.4 dB EPD. The threshold values are higher than the results presented in Fig. 6 of Rao et al. (2001), where they were on average 2.3 dB EPD. It could be possible that this difference in threshold is caused by a different amount of training. In the present experiment the subject with the longest training (S4) reached a mean threshold of 2.4 dB

EPD, which comes close to the grand mean found in (Rao et al., 2001). The thresholds of the other subjects might also be lower after a longer training period.

The overall differences between the subjects contribute to the variation in thresholds. For each individual subject, thresholds vary no more than about a factor 2 across conditions, indicating that the distance measure gives a reasonable indication of the subjects' sensitivity.

The next section presents a method to quantify the perceptual relevance of R -parameter changes, which is used to calculate directions of maximum, intermediate and minimum perceptual sensitivity. A second listening experiment, described in Section 6, is used to evaluate D for variations to the R -parameters in these directions.

5. Perceptual relevance of R -parameter changes

5.1. Quadratic approximation of the distance measure D

In this section an approximation of D is derived, which allows one to predict the directions of maximum and minimum perceptual relevance of a variation about the point r . In addition this method gives a measure for the minimum parameter change needed to produce a just noticeable difference (JND). This value will be called the perceptual relevance of the point r .

In order to quantify the perceptual relevance of R -parameter variations, the measure D will be approximated. In fact, we will approximate D^2 rather than D because we can assume that the partial derivatives of order two of D^2 exist and are continuous on a subset of \mathbb{R}^3 for all R -parameter vectors r . We can then express the Taylor expansion, consisting of a sum of relatively simple functions, of D_r^2 about a point r in terms of the gradient g_r , the Hessian H_r and the remainder term (or truncation error) $o(\|h\|^2)$:

$$D_r^2(h) = D_r^2(0) + g_r^T h + \frac{1}{2} h^T H_r h + o(\|h\|^2). \quad (4)$$

The Hessian H_r is a 3×3 matrix with elements $h_{ij} := \partial^2 D_r^2 / \partial r_i \partial r_j$. Because $D_r(0) = 0$ and

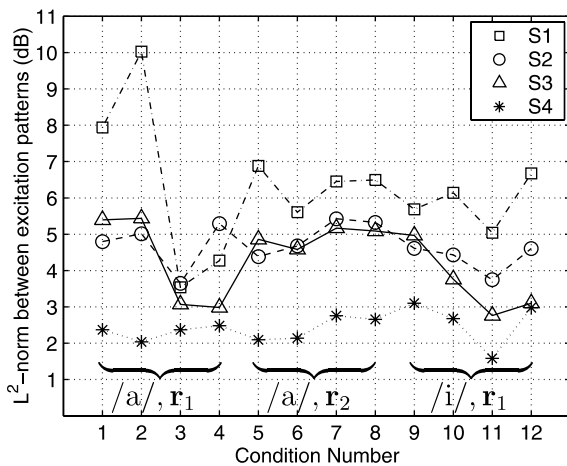


Fig. 4. Thresholds for the 12 conditions of Experiment I in terms of the excitation pattern distance. The results of the four subjects are indicated by the four different symbols.

$D_r^2(\mathbf{h}) \geq 0$ for all $\mathbf{h} \in \mathbb{R}^3$, $D_r^2(\mathbf{0})$ is a local minimum, thus the gradient \mathbf{g}_r^T has to be zero. This reduces Eq. (4) to

$$D_r^2(\mathbf{h}) = \frac{1}{2} \mathbf{h}^T H_r \mathbf{h} + o(\|\mathbf{h}\|^2). \quad (5)$$

Eq. (5) is approximated by omitting the remainder term $o(\|\mathbf{h}\|^2)$. The measure D is then approximated by the square root of a quadratic functional

$$Q_r(\mathbf{h}) := \sqrt{\frac{1}{2} \mathbf{h}^T H_r \mathbf{h}}. \quad (6)$$

Because the symmetric matrix H is positive definite, i.e. $\mathbf{h}^T H \mathbf{h} > 0$ for all $\mathbf{h} \in \mathbb{R}^3 \setminus \{\mathbf{0}\}$, the square root exists. Quadratic functionals are mathematically simple and therefore used in many applications of numerical analysis.

5.2. Quantifying the perceptual relevance

In order to quantify the perceptual relevance, we need to compute the eigenvalues and eigenvectors of the matrix H_r , where $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$ denote the eigenvalues of H_r and $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ the corresponding orthonormal eigenvectors. The square root of the eigenvalue λ_1 determines the perceptual effect for a variation of \mathbf{r} in the direction of \mathbf{v}_1 . If $\sqrt{\lambda_1}$ is small for a specific point \mathbf{r} , then a variation of \mathbf{r} in the direction \mathbf{v}_1 will only have a small perceptual effect, and since λ_1 is the largest eigenvalue, *all* variations of \mathbf{r} will have a small perceptual effect. Thus, $\sqrt{\lambda_1}$ can be used to determine the *overall perceptual sensitivity* of parameter variations about a point \mathbf{r} . The corresponding eigenvector \mathbf{v}_1 then determines the direction of the maximum perceptual sensitivity for a variation of \mathbf{r} . Likewise, the eigenvector \mathbf{v}_3 determines the direction of the minimum perceptual sensitivity.

Fig. 5 shows an example of a two-dimensional plot with axes R_1 and R_2 showing (hypothetical) contour lines of the quadratic functional $Q_r^2(\mathbf{h})$ for a specific point \mathbf{r} in the R -parameter space.⁴

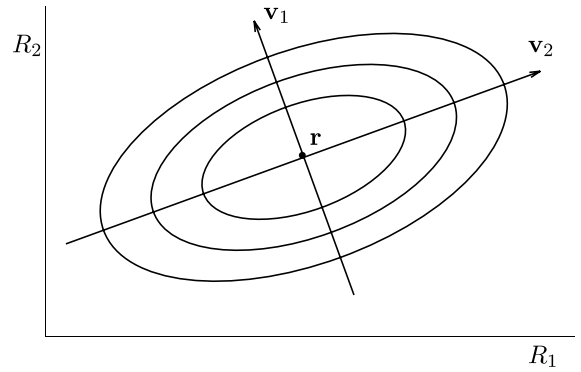


Fig. 5. Example of a two-dimensional plot showing contour lines of the quadratic functional $Q_r^2(\mathbf{h})$ for a point \mathbf{r} in the R -parameter space for small variations \mathbf{h} . The eigenvector \mathbf{v}_1 shows the direction of maximum sensitivity and \mathbf{v}_2 the direction of minimum sensitivity.

For all points $\mathbf{r} + \mathbf{h}$ on a contour line, the values of $Q_r^2(\mathbf{h})$ are constant and thus the perceptual distance from \mathbf{r} is the same. The directions of maximum and minimum perceptual sensitivity, \mathbf{v}_1 and \mathbf{v}_2 , are displayed in the figure.

To obtain one JND in the R -parameter space for a threshold change in the perceptual space, the inverse functions D^{-1} and Q^{-1} are used. The value of one JND in the R -parameter space can only be calculated when the direction in which the parameters are varied is known, because D and Q are mappings from \mathbb{R}^3 into \mathbb{R} . The following definitions will be used to calculate the inverse for a given direction in the R -parameter space. We will use the vector $\mathbf{u} \in \mathbb{R}^3$ with $\|\mathbf{u}\| = 1$ to indicate the direction of change and the nonnegative scalar τ to indicate the amount of change. For a given point \mathbf{r} we express the perceptual distance to a point $\mathbf{r} + \tau\mathbf{u}$ by defining the functions $D_{r,\mathbf{u}}(\tau) := D_r(\tau\mathbf{u})$ and $Q_{r,\mathbf{u}}(\tau) := Q_r(\tau\mathbf{u})$. For changes along the three eigenvectors \mathbf{v}_i , i.e. $\mathbf{u} = \mathbf{v}_i$, Q_{r,\mathbf{v}_i} can be expressed by the eigenvalues λ_i :

$$Q_{r,\mathbf{v}_i}(\tau) = Q_r(\tau\mathbf{v}_i) = \left(\frac{\lambda_i}{2}\right)^{1/2} \tau. \quad (7)$$

Since we need to compute the inverse functions D^{-1} and Q^{-1} , we have to check whether the functions D and Q are strictly increasing, at least in the

⁴ In the mathematical analysis, the problem is of course three-dimensional. For ease of visualisation, we have projected it to a two-dimensional subspace.

range of changes used in the experiments. This is indeed true for the measure D for all points \mathbf{r} and directions \mathbf{u} used in the first experiment. The value of τ corresponding to an EPD value of v can be obtained implicitly by determining τ in the equation $v = D_{\mathbf{r},\mathbf{u}}(\tau)$. The function is strictly increasing up to a value τ_{\max} , thus the value τ is unique. Because the equation is implicit, the calculation of τ may be tedious.

The inverse of Q , however, can be expressed explicitly. Because the function $Q_{\mathbf{r},\mathbf{u}}$ is strictly increasing on $[0, \infty)$, the inverse function $Q_{\mathbf{r},\mathbf{u}}^{-1}$ is defined on $[0, \infty)$. The values $Q_{\mathbf{r},\mathbf{u}}^{-1}(v)$ can be obtained by

$$Q_{\mathbf{r},\mathbf{u}}^{-1}(v) := \left(\frac{2}{\mathbf{u}^T H_{\mathbf{r},\mathbf{u}}} \right)^{1/2} v. \tag{8}$$

If we define the threshold in the perceptual space as v_0 , then differences $v \geq v_0$ are audible and differences $v < v_0$ are inaudible. The size of one JND in the direction of maximum perceptual sensitivity is $\sqrt{2/\lambda_1} \cdot v_0$. One JND in the R -parameter space for an arbitrary direction \mathbf{u} is given by $\tau_{\text{JND}} := D_{\mathbf{r},\mathbf{u}}^{-1}(v_0)$. In the same way we can compute one JND with respect to the functional Q : $\tilde{\tau}_{\text{JND}} := Q_{\mathbf{r},\mathbf{u}}^{-1}(v_0)$. Because a good approximation of D by Q will not immediately imply a good approximation of $D_{\mathbf{r},\mathbf{u}}^{-1}$ by $Q_{\mathbf{r},\mathbf{u}}^{-1}$, we examined the error between the inverse functions as well. For the computation of JNDs in the R -parameter space with the inverse of Q it is desired that the values τ_{JND} and $\tilde{\tau}_{\text{JND}}$ are close. In Section 7.2 we will see that $Q_{\mathbf{r},\mathbf{u}}^{-1}$ approximates actual JNDs rather well.

In the next section D is evaluated by means of a listening test for R -parameter variations in directions in maximum, intermediate and minimum perceptual sensitivity.

6. Experiment II: Parameter variations in directions chosen for perceptual relevance

In the second experiment, the R -parameters were varied in orthogonal directions \mathbf{v}_1 , \mathbf{v}_2 and \mathbf{v}_3 corresponding to the maximum, intermediate and minimum perceptual sensitivity, respectively, calculated with the method described in Section 5. This additional experiment was conducted to test the distance measure for these directions and to compare the results with Experiment I. The setup of the second experiment was similar to that of Experiment I. The variations were performed for three parameter sets $\mathbf{r}_3 := [0.092, 0.463, 0.791]^T$, $\mathbf{r}_4 := [0.028, 0.420, 0.660]^T$ and $\mathbf{r}_5 := [0.012, 0.374, 0.545]^T$ for the vowel /a/. The sets were chosen from a large collection of realistic data, and corresponded to small, intermediate and large values of λ_1 . The corresponding eigenvectors in directions of maximum perceptual relevance in which the three points were varied are $[-0.78, -0.57, 0.25]^T$, $[0.98, 0.18, -0.03]^T$ and $[-0.99, -0.13, 0.005]^T$, respectively. The entries of the vectors show the contributions of R_a , R_k and R_o , respectively.

As in Experiment I, the fundamental frequency was 110 Hz and the stimuli comprised the first 36 harmonics. Again, U_0 and E_c were allowed to vary with the R -parameters. In this experiment the signals always had regular phase. In total there were 9 conditions. In Table 2 the conditions of this experiment, numbered 13–21, are presented.

Four subjects (S2, S4, S5 and S6) participated in the experiments. All were young adults with normal hearing and no reported history of hearing impairment. Subjects S2 and S4 had participated in Experiment I. Subjects S5 and S6 were newly recruited because subjects S1 and S3 were no longer available. The stimuli were again presented

Table 2

Parameter conditions of Experiment II showing the directions of modifications to the R -parameters $\mathbf{r}_3 := [0.092, 0.463, 0.791]^T$, $\mathbf{r}_4 := [0.028, 0.420, 0.660]^T$ and $\mathbf{r}_5 := [0.012, 0.374, 0.545]^T$

Direction	Conditions and corresponding modifications		
	/a/, \mathbf{r}_3	/a/, \mathbf{r}_4	/a/, \mathbf{r}_5
\mathbf{v}_1	Condition 15	Condition 18	Condition 21
\mathbf{v}_2	Condition 14	Condition 17	Condition 20
\mathbf{v}_3	Condition 13	Condition 16	Condition 19

diotically over headphones at a level ranging from approximately 48–61 dB SPL.

6.1. Experiment II: Results

In Fig. 6 the JNDs of the subjects S2, S4, S5 and S6 are shown. On the y -axis a logarithmic scale is used. The diamonds and plus signs in the figures are the predicted JND values, calculated by the distance measures D and Q , which will be discussed in Section 7.1.

At first glance Fig. 6 shows different results than Fig. 3 in Section 4.2. A division into three magnitudes is not immediately evident. The data of the conditions corresponding to point r_3 are not as divergent as the data corresponding to the two other points r_4 and r_5 . In general one can see an ordering from left to right, where the parameter variations per point r_i become increasingly divergent.

This divergency can be related to the different locations of points in the R -parameter space at which the JNDs were measured. The R_a and R_o values of point $r_3 := [0.092, 0.463, 0.791]^T$ are much larger than the R_a and R_o values of the point

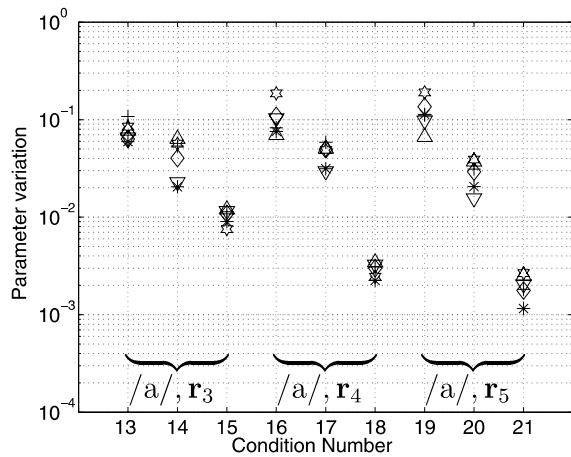


Fig. 6. Data obtained from Experiment II of the measured JNDs in the R -parameter space for each condition for points $r_3 := [0.092, 0.463, 0.791]^T$, $r_4 := [0.028, 0.420, 0.660]^T$ and $r_5 := [0.012, 0.374, 0.545]^T$. The diamonds indicate the predicted JND values $\tau_{i,JND}$ calculated with the inverse of D . The + signs indicate the predicted JND values $\tilde{\tau}_{i,JND}$ calculated with the inverse of Q . On the y -axes a logarithmic scale is used.

$r_5 := [0.012, 0.374, 0.545]^T$. The values of point $r_4 := [0.028, 0.420, 0.660]^T$ lie between the two points. As mentioned earlier, the points r_3 , r_4 and r_5 of Experiment II were chosen in such a way that the value of $\sqrt{\lambda_1}$ was small (536), intermediate (1960) and large (3338), resulting in a low, intermediate and high overall perceptual sensitivity to parameter variations about the points. This is reflected in the obtained data, where the ordering of $\sqrt{\lambda_1}$ of points r_3 , r_4 and r_5 corresponds to the ordering of the divergency of the JNDs per point in Fig. 6. The overall perceptual sensitivities of points r_1 and r_2 are 2390 and 3267, respectively, and are thus in the range of the overall perceptual sensitivity of points r_4 and r_5 . A comparison of Figs. 3 and 6 shows that the divergency of the points r_1 and r_2 corresponds to the divergency of points r_4 and r_5 .

In Fig. 7 the data from Fig. 6 are expressed in terms of the perceptual distance measure. The data of S2, S4, S5 and S6 are indicated with a circle, an asterisk, a star, and a downward triangle, respectively. It can be observed that the thresholds of the subjects for the R parameter changes lie between 2 and 7 dB EPD. The grand mean of the thresholds is 4.2 dB EPD. The data show that the ranges of the thresholds are somewhat closer than the ranges of the data from Experiment I, due to the fact that

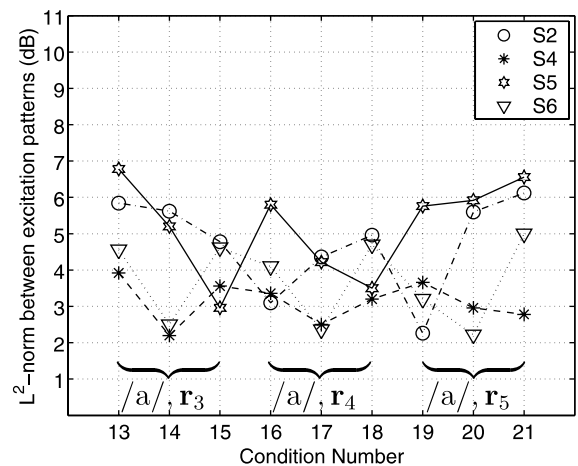


Fig. 7. Thresholds for the 9 conditions of Experiment II in terms of the excitation pattern distance. The results of the four subjects are indicated by the four different symbols.

in Experiment II the overall differences between subjects are smaller. In general, the results of this experiment are in terms of the EPD similar to the results of Experiment I.

7. Discussion of Experiments I and II

In this section the results of the two experiments are discussed and compared. We observe that the thresholds differ per subject. We can handle this observation in two ways. We can either look at the range of perceptual thresholds and JNDs in the R -parameter space across subjects or take one threshold and JND, by using the grand mean of the data. It is often desirable to maintain a single threshold value in the perceptual domain. The use of a single threshold value opens more possibilities for general, subject-independent research. If a single representative is not desired, the distance measure can be calibrated for each subject separately.

7.1. The choice of threshold and the size of one JND

To obtain a single threshold value, the grand mean of the thresholds is a reasonable choice. The grand means of Experiment I and II are 4.4 and 4.2 dB EPD, respectively. Because the grand means are rather close, we will take the total grand mean 4.3 dB EPD as the representative value.

With the inverse function of $D_{r,u}$ obtained in Section 5 we can determine the size of one JND by $\tau_{\text{JND}} := D_{r,u}^{-1}(4.3)$. The size of one JND in terms of the functional Q is denoted by $\tilde{\tau}_{\text{JND}} := Q_{r,u}^{-1}(4.3)$. Thus the symbols τ_{JND} and $\tilde{\tau}_{\text{JND}}$ represent predictions of the JNDs obtained from the inverse functions D and Q , respectively. In Section 7.2, τ_{JND} and $\tilde{\tau}_{\text{JND}}$ will be compared for the directions used in the two experiments.

For the 21 conditions $i \in \{1, \dots, 21\}$ and 6 subjects $j \in \{1, \dots, 6\}$ across the two experiments, a total of 84 thresholds $v_{i,j}$ in dB EPD (48 and 36 conditions for Experiment I and II, respectively) is available for this analysis. The individual JND values $\tau_{i,j}$ in the R -parameter space are shown in Figs. 3 and 6. The diamonds in the figures are the calculated values $\tau_{i,\text{JND}}$ for the conditions

$i \in \{1, \dots, 21\}$ obtained for the mean threshold of 4.3 dB EPD. The calculated values $\tilde{\tau}_{i,\text{JND}}$ are indicated with a plus sign in the figures. It can be observed that the values $\tau_{i,\text{JND}}$ are within the value range of the individual subjects. Thus, apparently JNDs can be estimated well by D^{-1} .

In the next subsection the results of Experiments I and II are used to quantify the error in the approximation of D by Q . If Q approximates D well, we can use this Q for the quantification of the perceptual relevance of the R -parameters.

7.2. Analysis of the accuracy of the quantification of the perceptual relevance

In this section the accuracy of the quantification of the perceptual relevance of the R -parameters is analysed. This accuracy depends on the error of the functional Q in representing the perceptual distance measure D . In this study it is desired that the functional Q approximates D well up to at least one JND in the R -parameter space. In addition, we will also investigate whether the inverse function Q^{-1} approximates D^{-1} well up to at least one JND, which is the case if $\tilde{\tau}_{i,\text{JND}}$ is close to $\tau_{i,\text{JND}}$ for $i \in \{1, \dots, 21\}$.

The functions D and Q were analysed using the results of the experiments in the perceptual space. The mean and standard deviation (std) of the subjects' thresholds $v_{i,j}$ in dB EPD calculated with the function D and the mean and std of the thresholds calculated with the function Q , denoted as $\tilde{v}_{i,j} := Q_r(\tau_{i,j})$ where $\tau_{i,j}$ are the measured JNDs, are compared. We found that the grand mean of $v_{i,j}$ is 4.3 and the std is 1.6 whereas the mean of $\tilde{v}_{i,j}$ is 4.4 and the std is 1.9. This outcome indicates that Q approximates D well. An additional t-test with a level of significance $\alpha = 0.05$ showed that the confidence interval for the mean for $\tilde{v}_{i,j}$ is (4.0, 4.8), thus including the selected mean value of 4.3 dB EPD.

We analysed the inverse functions with the results of the experiments in the R -parameter space. From Figs. 3 and 6 it can be observed that the values $\tau_{i,\text{JND}}$ and $\tilde{\tau}_{i,\text{JND}}$ are close to each other and almost for every condition within the range of the values $\tau_{i,j}$, except for condition 3 and 13 where $\tilde{\tau}_{3,\text{JND}}$ and $\tilde{\tau}_{13,\text{JND}}$ are just above the range of

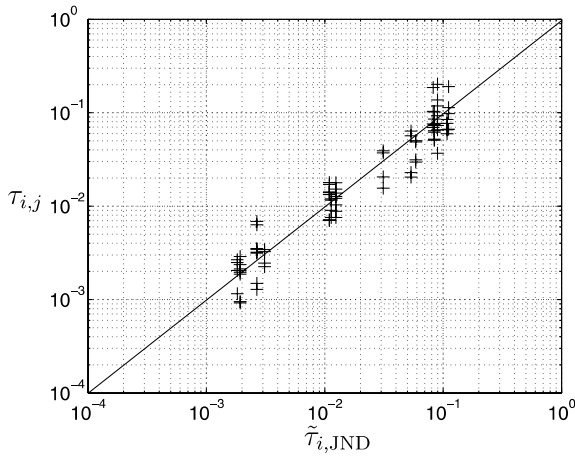


Fig. 8. Data of the obtained thresholds in terms of the distance between the points in the R -parameter space calculated with the inverse of D and Q . The values $\tau_{i,j}$ are plotted as function of $\tilde{\tau}_{i,j,JND}$ for the conditions of both experiments. On both axes logarithmic scales are used.

experimentally obtained JNDs. In Fig. 8 the values $\tilde{\tau}_{i,JND}$ (x -axis) are plotted against the values $\tau_{i,j}$ (y -axis) for all conditions and measured thresholds $v_{i,j}$ in dB EPD. The figure shows that the values lie around the 45° line. From this observation and from Figs. 3 and 6 one can conclude that the inverse $Q_{r,u}^{-1}$ behaves relatively well up to at least one JND in the R -parameter space.

7.3. Comparison with other results

Scherer et al. (1998) measured JNDs for variations of the parameters S_q and O_q for one specific parameter setting ($S_q = 2$ and $O_q = 0.6$) of the glottal-pulse. The JNDs were measured for the signals of the glottal-flow alone and for the vowel waveforms at an overall level of 74 dB(A) for all stimuli. In contrast to our experiments, the flow amplitude U_0 of the glottal-flow was held constant for all the realised glottal-pulses. The parameters S_q and O_q are closely related to the R -parameters, with $R_k := 1/S_q$ and $R_o := O_q$. For comparison we expressed the parameters and JNDs of Scherer et al. in terms of the R parameters. The average JND for a variation of the parameter R_o was 0.0344, the average JND for a variation in the direction R_k was 0.0688 for synthesised signals of

vowel waveforms. In Experiment I of this paper the JNDs obtained for R_o were 0.085 (average JNDs of conditions 3, 4, 11 and 12) and for R_k , they were 0.012 (average JNDs of conditions 7, 8, 9 and 10). The JNDs for R_o obtained by Scherer et al. are a little bit below the range of the JNDs obtain from our experiment, while the JNDs for R_k are almost a factor 6 higher than the average of the JNDs from our experiment. The reason for this discrepancy lies in the fact that variations of the parameter R_k have a strong effect on the flow amplitude of the signal as well as on the maximum excitation E_e and thus on the overall level of the signal. In the experiment of Scherer et al. with constant flow amplitude of the stimuli, the R_k -parameter had to be varied much more to reach an audible difference compared with our experiments where the flow amplitude was not held constant.

Because the R_a values for Scherer's data were unknown, we computed the EPDs for the measured JNDs at different values of R_a ranging from 0 to 0.15. For JNDs measured in the direction of R_k the EPDs lie between 5.8 and 8.9 dB EPD, and in the direction of R_o between 3.2 and 14.7 dB EPD.

Another way of looking at Scherer's results is by computing EPDs from the spectra shown in Figs. 4 and 9 in (Scherer et al., 1998). We calculated the EPDs between the spectra of the standard and of the stimuli corresponding to the measured JNDs for R_o and R_k . For variations of R_o and R_k the values were 8.2 and 9.0 dB EPD, respectively. The calculated EPDs are higher than the obtained thresholds in our experiments. Apart from the higher thresholds, it can be observed that the calculated EPDs vary no more than a factor 1.1. This shows that the thresholds are rather constant over the two varied directions R_o and R_k . If we compute the EPDs for the JNDs measured by Scherer with an R_a value of 0.072 then the EPDs correspond to the values 8.2 and 9.0 dB EPD measured from the spectra. Campbell et al. (1998) also examined the relation between glottal-pulse parameters and spectral and perceptual distance measures for the data of Scherer et al. (1998). They found that spectral distance measures and perceptual distance measures performed equally well for these data. We compared the Euclidean distance between ex-

citation patterns used by Campbell et al. with our EPD. It appears that another type of distance measure was used by these authors, for the thresholds for variations of R_o and R_k differed about a factor 2, while with the EPD they differ only by a factor 1.1.

Henrich et al. (in press) measured JNDs for three values of the O_q (0.4, 0.6 and 0.8) and for two values of the asymmetry coefficient α_m (2/3 and 0.8). The values of α_m can be transformed into values of R_k by using the following relation: $R_k := (1 - \alpha_m)/\alpha_m$. The time interval of the return phase was chosen to be zero, hence $R_a = 0$. In all but one of their conditions, the flow amplitude U_0 was held constant. In the remaining condition the parameter E_c was held constant. The JNDs were measured for the vowels /a/ and /i/ and the fundamental frequencies 130 and 196 Hz. The effect of vibrato was also tested. It was found that when the amplitude parameter E_c was held constant the measured JNDs increased significantly. Also the location from which a parameter setting was varied affected the measured JNDs. The vibrato did have a slight effect on the JNDs, while funda-

mental frequency and choice of vowel did not have any influence on the measured JNDs.

A group of 20 untrained subjects and a group of 10 trained subjects participated in their experiments. The JNDs for the three R_o values, averaged across the results of the untrained subjects, were between 0.058 and 0.106 for untrained subjects and between 0.037 and 0.079 for the trained subjects. The JNDs for the parameter α_m , expressed in terms of R_k were between 0.041 and 0.071 for untrained subjects and between 0.034 and 0.058 for trained subjects. When the amplitude parameter E_c was held constant the JNDs were larger (see Table 3). By using the measured JND values of Henrich et al., we computed the distances in the perceptual domain to compare their results with ours. The JND data of Henrich et al. are shown in Table 3 together with the EPDs computed with our distance measure. In the computation of the EPDs we did not add vibrato to the stimuli.

We see that the ranges of the computed EPDs of both groups are very narrow. The thresholds of the untrained subjects are somewhat higher than

Table 3

Average JNDs of Henrich et al. (in press) for untrained and trained subjects with corresponding computed EPDs

Var.	R_o	R_k	F_0 [Hz]	Vowel	U_0 constant		E_c constant	
					JNDs	EPDs	JNDs	EPDs
R_o	0.40	0.50	196	/a/	0.058	6.9		
* R_o	0.60	0.50	196	/a/	0.068	6.6		
R_o	0.60	0.50	196	/a/	0.087	8.0	0.173	7.6
R_o	0.60	0.50	130	/a/	0.092	7.5		
R_o	0.80	0.50	196	/a/	0.106	7.2		
R_k	0.60	0.50	196	/a/	0.071	7.9		
R_k	0.60	0.25	196	/a/	0.041	6.7		
R_o	0.60	0.50	196	/i/	0.080	7.0		
R_o	0.40	0.50	196	/a/	0.037	4.7		
* R_o	0.60	0.50	196	/a/	0.041	4.1		
R_o	0.60	0.50	196	/a/	0.063	6.1	0.099	5.0
R_o	0.60	0.50	130	/a/	0.071	6.0		
R_o	0.80	0.50	196	/a/	0.079	5.6		
R_k	0.60	0.50	196	/a/	0.058	6.5		
R_k	0.60	0.25	196	/a/	0.034	5.5		
R_o	0.60	0.50	196	/i/	0.053	4.8		

The first five columns indicate the parameters of the experiment in terms of which parameter was varied (column 1), the values of R_o and R_k for the reference stimulus (columns 2 and 3) and F_0 and the chosen vowel (columns 4 and 5). The measured JNDs are expressed here in terms of R_o and R_k . The top half shows the results of the untrained subjects and the bottom half the results of the trained subjects. For the conditions marked with a star, no vibrato was used in the experiment.

the thresholds of the trained subjects, but this learning effect reduces the EPD value only by 2 dB. In the table we see that the JNDs can vary largely, but the EPDs stay in a very close range. It can be stated that these results are in good agreement with the results from our own study. The EPDs are just slightly higher than 4.3 dB EPD, probably due to the presence of vibrato which can be observed from the table, and lie in the range of the measured EPDs from our experiments. The outcome of this analysis supports the value and applicability of the proposed perceptual distance measure.

8. Conclusions

A method for the analysis of the perceptual relevance of glottal-pulse parameter variations was developed. The perceptual distance measure D based on excitation patterns was evaluated in order to predict audibility discrimination thresholds for small changes to the R parameters of the LF model. Listening tests were performed to validate the distance measure D and to determine the size of one just noticeable difference in the R -parameter space. The distance measure D was approximated by the square root of a quadratic functional. The properties of the quadratic functional were used to determine the directions of maximum/minimum sensitivity. From our experiments and model analysis we observed that the measured JNDs strongly depend on the values of the parameters in the R -parameter space from where the R parameters are varied. The overall perceptual effect of variations of R -parameters can be expressed by the perceptual parameter $\sqrt{\lambda_1}$, which is the square root of the largest eigenvalue obtained from the approximation Q .

It was shown that the quadratic functional is a good approximation of D^2 for distances up to one JND, implying that the method can be used to quantify the perceptual relevance of the R -parameters. Next, the inverse functions of D and Q were compared in the R -parameter space at threshold levels of the subjects. The threshold values $\tilde{\tau}_{i,\text{JND}}$ in dB EPD calculated with the function Q are close to $\tau_{i,\text{JND}}$ in dB EPD calculated with the function D and lie almost everywhere within

the ranges of the subjects' measured JNDs $\tau_{i,j}$. It can be concluded that the inverse function of Q can be used to determine JNDs in the R -parameter space.

There are a number of applications of the method developed in this paper. One example is the possibility to divide the R -parameter space into cells within which sets of R -parameters are perceptually indistinguishable, which can be used in speech synthesis. The method also provides a way to quantify systematically the relation between glottal-pulse parameters and voice quality. The application of our method to R -parameter vectors taken from literature will give a better understanding of the R -parameter space. In (van Dinter et al., 2001) it was already found that the R -parameter vectors can be ordered along a trajectory which is a function of the overall perceptual relevance $\sqrt{\lambda_1}$. This interesting result will be presented and discussed in an extended follow-up paper.

References

- Campbell, S., Arehart, K., Scherer, R., Butler, A., 1998. Application of auditory models to discrimination thresholds of voicing parameters. In: Proc. Joint Meeting of the 16th Internat. Congress on Acoustics and the 135th Meeting of the Acoustical Society of America, Vol. 4, pp. 2603–2604.
- Childers, D., Lee, C., 1991. Voice quality factors: analysis, synthesis and perception. *J. Acoust. Soc. Amer.* 90, 2394–2410.
- Doval, B., d'Alessandro, C., 1997. Spectral correlates of glottal waveform models: an analytic study. In: Proc. ICASSP-97, pp. 1295–1298.
- Doval, B., d'Alessandro, C., Diard, B., 1997. Spectral methods for voice source parameters estimation. In: Proc. EURO-SPEECH'97, Vol. 1, pp. 533–536.
- Fant, G., Liljencrants, J., Lin, Q., 1985. A four-parameter model of glottal flow. In: STL-QPSR, Vol. 4/85, KTH, pp. 1–3.
- Gobl, C., 1989. A preliminary study of acoustic voice quality correlates. In: STL-QPSR, Vol. 4/89, pp. 9–22.
- Gobl, C., Ní Chasaide, A., 1992. Acoustic characteristics of voice quality. *Speech Comm.* 11, 481–490.
- Henrich, N., Sundlin, G., Ambroise, D., d'Alessandro, C., Castellengo, M., Doval, B., 2003. Just noticeable differences of open quotient and asymmetry coefficient in singing voice. *J. Voice*, in press.
- Karlsson, I., 1990. Voice source dynamics for female speakers. In: Proc. ICSLP'90, Kobe, Japan, pp. 69–72.

- Karlsson, I., Liljencrants, J., 1996. Diverse voice qualities: models and data. In: *TMH/QPSR*, Vol. 2/96, KTH, pp. 143–146.
- Klatt, D., Klatt, L., 1990. Analysis synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Amer.* 87, 820–856.
- Levitt, H., 1971. Transformed up–down methods in psychoacoustics. *J. Acoust. Soc. Amer.* 49, 971–995.
- Moore, B., 1986. *Frequency Selectivity in Hearing*. Academic Press Inc., London.
- Moore, B., Glasberg, B., Baer, T., 1997. A model for the prediction of thresholds, loudness and partial loudness. *J. Audio Eng. Soc.* 45, 224–240.
- Plomp, R., Steeneken, H., 1969. Effect of phase on the timbre of complex tones. *J. Acoust. Soc. Amer.* 46, 409–421.
- Rao, P., van Dinther, R., Veldhuis, R., Kohlrausch, A., 2001. A measure for predicting audibility discrimination thresholds for spectral envelope distortions in vowel sounds. *J. Acoust. Soc. Amer.* 109, 2085–2097.
- Scherer, R., Arehart, K., Guo, C., Milstein, C., Horii, Y., 1998. Just noticeable differences for glottal flow waveform characteristics. *J. Voice* 12, 21–30.
- van Dinther, R., Kohlrausch, A., Veldhuis, R., 2001. The perceptual relevance of glottal-pulse parameter variations. In: *Proc. EUROSPEECH'01*, Vol. 2, pp. 1507–1510.
- Veldhuis, R., 1998. The spectral relevance of glottal-pulse parameters. In: *Proc. ICASSP*, Seattle, pp. 873–876.
- Zwicker, E., Scharf, B., 1965. A model of loudness summation. *Psychol. Rev.* 72, 3–26.