Available online at www.sciencedirect.com
SCIENCE DIRECT®
Speech Communication 46 (2005) 95–112
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Perceptual aspects of glottal-pulse parameter variations

R. van Dinther [a], R. Veldhuis [b], A. Kohlrausch [c,d,*]

[a] Department of Physiology, Centre for the Neural Basis of Hearing, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK
[b] Faculty of Engineering, Chair of Signals and Systems, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[c] Department of Technology Management, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[d] Philips Research Laboratories Eindhoven, Prof. Holstlaan 4, WO 02, 5656 AA Eindhoven, The Netherlands

Received 17 June 2004; received in revised form 4 January 2005; accepted 11 January 2005

Abstract

The relation between speech production and perception parameters is investigated for synthetic stationary vowels. A perceptual distance measure, based on the distance between excitation patterns, is used to quantify and model the perceptual relevance of variations to glottal-pulse parameters. The parameters that are studied are the *R* parameters of the well-known Liljencrants–Fant (LF) model. An approximation of the perceptual distance measure was used to quantify the perceptual effects of the *R* parameters. The data used in this paper consist of 33 *R*-parameter vectors which all have been measured from real voices and have been taken from the literature. The 33 points in the *R*-parameter space are found to lie near a trajectory when ordered as function of a perceptual parameter, which was derived from the perceptual distance measure. This ordering seems to be largely independent of other speech parameters, such as F_0 values, level and of parameters of the vocal tract filters. Finally, it is demonstrated that the perceptual parameter has a close relation to the production parameter R_d, described in [Fant, G., 1995. The LF-model revisited. Transformations and frequency domain analysis. STL-QPSR 2–3/95, 119–156].

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

1. Introduction

Glottal-pulse models have been of interest for quite some time. On the one hand, they provide insight in human speech production. On the other hand, they can help to increase the quality of synthesised speech (Klatt and Klatt, 1990). Much attention has been paid to relations between the

* Corresponding author. Address: Philips Research Laboratories Eindhoven, Prof. Holstlaan 4, WO 02, 5656 AA Eindhoven, The Netherlands. Tel.: +31402743093; fax: +31402744675.
E-mail address: armin.kohlrausch@philips.com (A. Kohlrausch).

0167-6393/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2005.01.005

parameters of these models and voice quality. According to Laver (1980), the definition of voice quality is the characteristic auditory colouring of an individual speaker's voice. The focus of this definition is an auditory phenomenon. This paper adopts Laver's definition and, therefore, considers voice quality as a perceptual characteristic. On the other hand, voice quality is a function of the vocal system, consisting of the lungs, larynx and supralaryngeal vocal tract. This means that the parameters of this vocal system directly influence voice quality. Consequently, studying the perceptual aspects of vocal-system parameters will contribute to a better understanding of voice quality. This paper deals with the perceptual relevance of small variations applied to speech production parameters, which has received much less scientific attention.

In speech synthesis, control of voice quality will improve the naturalness of synthetic speech. Modern speech synthesis is mainly concatenative synthesis, i.e. speech segments, derived from recorded speech, are retrieved from a database and concatenated. The segment length may be small and fixed as is the case in diphone synthesis, or it may be arbitrary, as in unit-selection based synthesis. There exist several techniques for waveform manipulation and morphing (McAulay and Quatieri, 1986; Moulines and Charpentier, 1990; Kawahara et al., 1999; Kawahara and Matsui, 2003) considering the modification of rhythm, pitch and emotional speech. The approach used in this study may help to develop techniques to modify voice quality of recorded speech in a systematic way.

In this paper the Liljencrants–Fant (LF) model (Fant et al., 1985) is adopted for the investigation of the perceptual relevance of small variations to glottal-pulse parameters. The LF model characterises the shape of the glottal pulse 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 perceptual relevance of small variations to the R parameters of the LF model is measured through the changes in the excitation pattern (Moore et al., 1997). A method based on an approximation of the distance between excitation patterns was developed in (Van Dinther et al., 2004) which allows to compute the perceptual relevance of small changes to the R parameters. In earlier studies (Scherer et al., 1998; Henrich et al., 2003), the perceptual relevance of speech production parameters was measured in the directions of one parameter only. The method described by Van Dinther et al. (2004) allows to quantify the perceptual relevance of variations in arbitrary directions in the parameter space. Furthermore, this method provides a way to determine the directions of maximum and minimum perceptual relevance. A perceptual parameter ψ was derived from this method to quantify the overall perceptual sensitivity of parameter variations about a parameter setting. In this paper, this perceptual parameter ψ is used to quantify the perceptual effects of R parameters measured from real voices which have been taken from the literature. Next, the relation between the R parameters and this perceptual parameter ψ is investigated.

A total of 33 R-parameter vectors derived from various publications, presenting a variety of voice qualities, was investigated. All vectors have been measured from recorded vowels. We show that the R parameters in the vectors can be approximated as simple functions of the perceptual parameter ψ, which means that all the parameter vectors are points near one trajectory in the 3-dimensional R-parameter space. Fant (1995) explored systematic covariations in the R parameters in a transformed version, in which characteristic trends are quantified as functions of a single shape parameter R_d. It turns out that this production parameter R_d is closely related to the perceptual parameter ψ. Fant (1995) indicated that there exists a close relation between voice qualities and the parameter R_d. On the basis of this and because of the simple relation between the parameters R_d and ψ we conclude that the perceptual parameter ψ can also be linked to voice quality.

The perceptual effects of the R-parameter variations can only be analysed from the utterance of the speech signal. This implies that the perceptual relevance of small variations to speech production parameters may depend on other parameters, such as the fundamental frequency F_0, level and the parameters of the vocal-tract filter. In this paper

we show that the perceptual parameter ψ is robust for different F_0 values, formant frequencies and bandwidths and for different overall levels of the signal.

The outline of this paper is as follows. Section 2 briefly discusses the LF model and the method used to quantify the perceptual relevance of small variations to the R parameters. The perceptual effects of such variations are studied in Section 3 for a number of vectors of estimated glottal-pulse parameters that were taken from literature. This section includes an analysis of the perceptual effects of R-parameter variations which are compared for different speech settings. In Section 4 the trajectory derived from the R parameters is examined more thoroughly. Section 5 compares Fant's shape parameter R_d with the perceptual parameter ψ. Section 6 presents the conclusions.

2. Quantifying the perceptual relevance of glottal-pulse parameter variations

The LF model has become a standard model for glottal-pulse analysis. Fig. 1 shows, as an 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 parameters T_0, T_p, T_e and T_a are called the T-parameters. The maximum airflow of the glottal pulse, often denoted as U_0, occurs at T_p and the maximum excitation with amplitude E_e occurs at T_e. In this paper the related set of R parameters is used, 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 effects of these shape parameters are represented in this paper in the following way: $\boldsymbol{r} := [R_a, R_k, R_o]^T \in \mathbb{R}^3$. The set of points $\boldsymbol{r}$ will be denoted as the R-parameter space $\mathscr{R}$ which is a subset of $\mathbb{R}^3$. The parameter space is not the entire $\mathbb{R}^3$. To get representative glottal-pulse waveforms from the R-parameter settings, the R-parameter space $\mathscr{R}$ is bounded by the following inequalities: $0 < R_a < 1 - R_o$; $0 < R_k < 1$; and $0 < R_o < 1$.

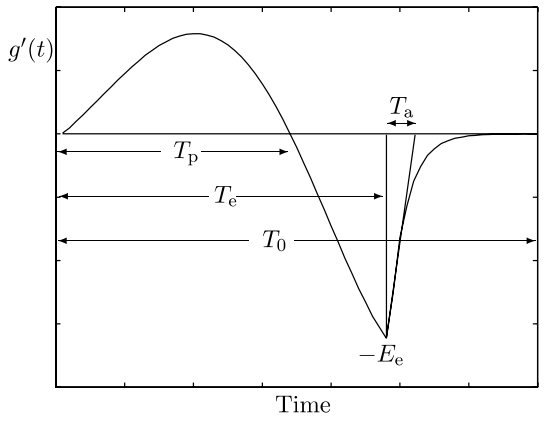

Fig. 1. Time derivative of one period of the glottal flow and definition of glottal-pulse parameters.

In order to quantify the perceptual effects of small variations to the R parameters, a perceptual distance measure is required. This distance measure is based on a mapping from the R-parameter space into a perceptual space. As elements of the perceptual space, representations of a sound in the inner ear are used, called excitation patterns. According to Moore (1987), excitation patterns can be thought of as the distribution of "excitation" evoked by a particular sound in the inner ear along a frequency axis. The excitation patterns are expressed in dB and are computed according to a model for computing loudness and excitation patterns (Moore et al., 1997). The excitation pattern is in this paper presented on a scale derived from the ERB width 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. An increase in frequency corresponding to the ERB value at that frequency represents a step of one on the ERB scale (Moore, 2003). Thus, the auditory filters are uniformly spaced on this ERB scale. In this paper, the excitation patterns are calculated for ERB scale values between 0 and 40, which corresponds to a frequency range of about 0–15 kHz. The notation $e(\boldsymbol{r}; x)$ is used to indicate the amount of excitation, derived for the parameter vector $\boldsymbol{r}$, as a function of the ERB scale x.

For a point $\boldsymbol{r}$ in the R-parameter space and $x \in (0, 40)$ the excitation pattern $e(\boldsymbol{r}; x)$ is calculated from the power spectrum of a 300 ms section of the time signal, derived from a source-filter model consisting of a combination of a source producing the

glottal-pulse derivative according to the LF model and a filter modelling vowel formants. The distance between the two excitation patterns, also called the excitation pattern distance (EPD), corresponding to a R-parameter vector $\boldsymbol{r}$ and a small change $\boldsymbol{r}+\boldsymbol{h}$ in the direction $\boldsymbol{h}$ is then quantified by

$$D_{\boldsymbol{r}}(\boldsymbol{h}) := \left(\int_0^{40} |e(\boldsymbol{r}+\boldsymbol{h};x) - e(\boldsymbol{r};x)|^2 \,\mathrm{d}x\right)^{\frac{1}{2}}. \tag{1}$$

The perceptual distance between two sounds will be denoted as the quantity dB EPD. The effectiveness of this distance for measuring discrimination thresholds was evaluated in (Rao et al., 2001). In order to quantify the perceptual relevance of R-parameter variations, $D_{\boldsymbol{r}}^2$ is approximated by means of its Taylor expansion. The Taylor expansion of $D_{\boldsymbol{r}}^2$ about a point $\boldsymbol{r}$ can be expressed in terms of the gradient $\boldsymbol{g}_{\boldsymbol{r}}$, the 3×3 matrix $H_{\boldsymbol{r}}$ with elements $h_{ij} := \partial^2 D_{\boldsymbol{r}}^2/\partial r_i \partial r_j$ and the remainder term $o(\|\boldsymbol{h}\|^2)$,

$$D_{\boldsymbol{r}}^2(\boldsymbol{h}) = D_{\boldsymbol{r}}^2(\boldsymbol{0}) + \boldsymbol{g}_{\boldsymbol{r}}^{\mathrm{T}}\boldsymbol{h} + \frac{1}{2}\boldsymbol{h}^{\mathrm{T}} H_{\boldsymbol{r}} \boldsymbol{h} + \mathrm{o}(\|\boldsymbol{h}\|^2). \tag{2}$$

Because $D_{\boldsymbol{r}}(0) = 0$ and $D_{\boldsymbol{r}}^2(\boldsymbol{h}) \geqslant \boldsymbol{0}$ for all $\boldsymbol{h} \in \mathbb{R}^3$, $D_{\boldsymbol{r}}^2(0)$ is a local minimum, thus the gradient $\boldsymbol{g}_{\boldsymbol{r}}^{\mathrm{T}}$ has to be zero. This reduces Eq. (2) to

$$D_{\boldsymbol{r}}^2(\boldsymbol{h}) = \frac{1}{2}\boldsymbol{h}^{\mathrm{T}} H_{\boldsymbol{r}} \boldsymbol{h} + \mathrm{o}(\|\boldsymbol{h}\|^2). \tag{3}$$

The symmetric matrix $H_{\boldsymbol{r}}$ is positive definite, i.e. $\boldsymbol{x}^{\mathrm{T}} H_{\boldsymbol{r}} \boldsymbol{x} > 0$ for all $\boldsymbol{x} \in \mathbb{R}^3 \setminus \{0\}$, with real eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 > 0$ and corresponding orthonormal eigenvectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$. The matrix $H_{\boldsymbol{r}}$ depends on $\boldsymbol{r}$ as well as on other speech parameters, such as the value of F_0 and the realised vowel. By omitting the remainder term $\mathrm{o}(\|\boldsymbol{h}\|^2)$, the measure $D_{\boldsymbol{r}}$ can be approximated by the square root of a quadratic functional

$$Q_{\boldsymbol{r}}(\boldsymbol{h}) := \sqrt{\frac{1}{2}\boldsymbol{h}^{\mathrm{T}} H_{\boldsymbol{r}} \boldsymbol{h}}. \tag{4}$$

For a more elaborate description of the measure $Q_{\boldsymbol{r}}$ the reader is referred to Van Dinther et al. (2004). Since the vector $\boldsymbol{v}_1$ belongs to the largest eigenvalue λ_1, changes in the direction v_1 lead to the strongest changes in the excitation pattern. The measure $Q_{\boldsymbol{r}}$ is therefore most sensitive to parameter variations along the direction $\boldsymbol{v}_1$. The value of λ_1 quantifies this sensitivity and is a direct measure for the perceptual strength of a change along the direction $\boldsymbol{v}_1$. For larger values of λ_1 smaller changes in the direction $\boldsymbol{v}_1$ will result in an audible difference. Because $\boldsymbol{v}_1$ indicates the direction of maximum perceptual relevance, λ_1 is a measure of the overall perceptual relevance for small variations to a specific point in $\mathscr{R}$. Actually, because of the square root in Equation (4), the square root of the eigenvalue is the perceptual parameter we will refer to in this paper. From now on the overall perceptual relevance of small variations to a specific point $\boldsymbol{r} \in \mathscr{R}$ will be denoted as

$$\psi := \sqrt{\lambda_1}. \tag{5}$$

This perceptual parameter will play an important role in this paper and it will be shown that it has an unforeseen direct relation with a speech-production parameter.

It was demonstrated in (Van Dinther et al., 2004) that $D_{\boldsymbol{r}}$ as well as its approximation $Q_{\boldsymbol{r}}$ are appropriate to predict perceptual discrimination thresholds resulting from R-parameter variations. It was additionally shown that the inverse of the function $Q_{\boldsymbol{r}}$ can be used to determine the change in the R-parameter space $\mathscr{R}$ corresponding to a just noticeable difference (JND) in the perceptual space. If the threshold in the perceptual space is defined as v_0, then differences $v \geqslant v_0$ are audible and differences $v < v_0$ are inaudible. The size of one JND in the direction of maximum perceptual sensitivity $\boldsymbol{v}_1$ is

$$v_0\sqrt{2}/\psi. \tag{6}$$

It was found that the threshold v_0 for R-parameter variations is about 4.3 dB EPD.

In the next section the overall perceptual relevance of small variations to R parameters is calculated for 33 R-parameter vectors and the relation of the parameter vectors with ψ is discussed.

3. The perceptual effects of the *R* parameters

3.1. A relation between speech production and perception

As mentioned in Section 1, voice quality is considered as a perceptual characteristic. It is

therefore argued that a classification of voice qualities can be found through a perceptual parameter. This can be achieved systematically by finding a relation between speech production parameters, which are related to voice quality, and a perceptual parameter.

The speech production parameters were taken from the references Childers and Lee (1991) and Karlsson and Liljencrants (1996), which were measured from utterances of different voice qualities. Table 1 shows the R-parameter vectors with the corresponding voice qualities. The R parameters 1–9 are taken from (Childers and Lee, 1991), the parameters 10–27 from (Karlsson and Liljencrants, 1996) and the parameters 28–33 are obtained from own research. This set is called W_r, where the subscript "r" stands for "real R-parameter vectors".

We computed the eigenvalues and eigenvectors of H_r for the 33 R-parameter vectors. The speech parameters were the vowel /a/, $F_0 = 110$ Hz, with formant frequencies $\{790, 1320, 2340, 3600\}$ [Hz] and corresponding bandwidths $\{90, 90, 142, 210\}$ [Hz]. In Fig. 2 the results are plotted for the set of 33 R-parameter vectors. The top panel shows the absolute values of the entries $v_{1,i}$ of the eigenvector $v_1 := [v_{1,1}, v_{1,2}, v_{1,3}]$ as a function of ψ. The values of the entries $v_{1,1}$, $v_{1,2}$ and $v_{1,3}$ correspond

Table 1
The 33 R-parameter vectors with the corresponding F_0 values and voice qualities obtained from various papers

	R_a [10^{-2}]	R_k [10^{-2}]	R_o [10^{-2}]	F_0 [Hz]	Voice quality
1	2.1	30.6	64.0	106	Modal
2	2.5	34.0	71.0	127	Modal
3	1.5	33.3	68.0	154	Modal
4	0.8	28.6	63.0	84	Slight vocal fry
5	0.5	25.0	25.0	45	Vocal fry
6	13.3	35.1	77.0	344	Falsetto
7	4.3	43.6	89.0	213	Falsetto
8	6.8	41.7	68.0	137	Breathy
9	10.0	44.8	84.0	200	Breathy
10	2.0	37.7	54.0	126	Normal (male)
11	5.0	51.7	71.0	246	Normal (female)
12	2.6	42.5	61.0	102	Low F_0 (male)
13	5.1	41.9	76.0	190	Low F_0 (female)
14	1.5	45.0	56.0	131	Medium F_0 (male)
15	4.2	48.0	71.0	250	Medium F_0 (female)
16	9.9	32.1	87.0	288	High F_0 (male)
17	3.0	48.7	65.0	360	High F_0 (female)
18	2.7	40.7	69.0	129	Low level (male)
19	10.5	57.1	81.0	249	Low level (female)
20	1.9	45.0	57.0	127	Medium level (male)
21	3.7	51.2	68.0	258	Medium level (female)
22	1.6	37.7	49.0	132	High level (male)
23	1.9	52.2	64.0	257	High level (female)
24	4.6	51.0	65.0	131	Breathy (male)
25	8.1	48.3	79.0	254	Breathy (female)
26	1.3	39.5	41.0	128	Pressed (male)
27	3.2	49.9	71.0	261	Pressed (female)
28	0.3	30.0	52.0	110	Tense
29	0.6	50.0	69.0	110	Modal
30	2.0	51.0	82.0	110	Lax
31	1.1	25.0	41.0	110	Tense
32	1.8	37.0	54.0	110	Modal
33	3.5	43.0	65.0	110	Lax

The R parameters 1–9 are taken from (Childers and Lee, 1991), the parameters 10–27 from (Karlsson and Liljencrants, 1996) and the parameters 28–33 are obtained from own research.

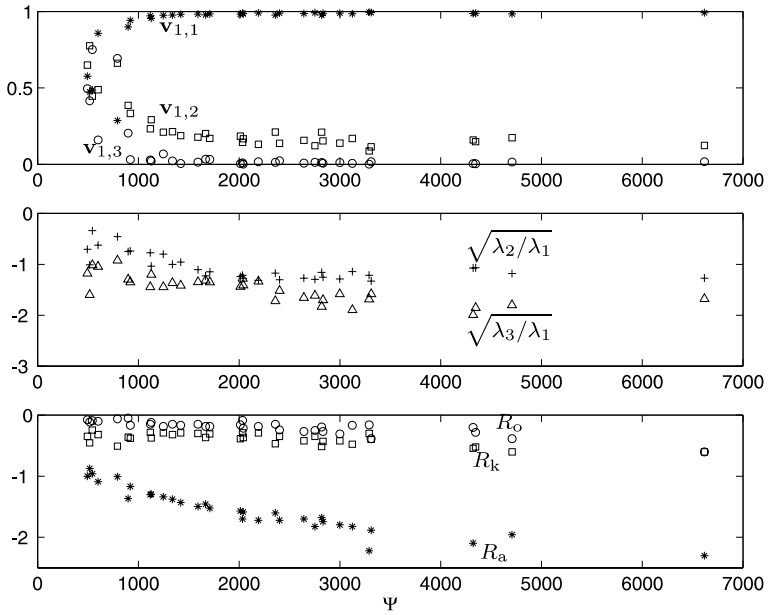

Fig. 2. Results for the vowel /a/, F_0 = 110 Hz for the R-parameter set W_r. Top panel: Absolute values of the entries of $\boldsymbol{v}_1$ as functions of ψ; the symbols star, square and circle correspond to the entries $v_{1,1}$, $v_{1,2}$ and $v_{1,3}$, respectively. Middle panel: Base-10 logarithms of $\sqrt{\lambda_2/\lambda_1}$ (plus signs) and $\sqrt{\lambda_3/\lambda_1}$ (upward triangles) as functions of ψ. Bottom panel: Base-10 logarithms of the R parameters as functions of ψ; R_o (circles), R_k (squares) and R_a (stars).

to the contributions of R_a, R_k and R_o, respectively. The middle panel shows the base-10 logarithms of the square-root eigenvalue ratios $\sqrt{\lambda_2/\lambda_1}$ and $\sqrt{\lambda_3/\lambda_1}$ as function of ψ, to compare the perceptual effects of small R-parameter variations in the directions $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ with effects of variations in the perceptually most significant direction $\boldsymbol{v}_1$. The bottom panel shows the base-10 logarithms of the R parameters as function of ψ. The perceptual parameter ψ orders the elements of the set W_r.

The top panel of the figure shows that for, approximately, $\psi > 1000$, the $\boldsymbol{v}_1$ direction is nearly constant and corresponds to a variation mostly in the R_a ($v_{1,1}$) and slightly in the R_k direction ($v_{1,2}$). The contribution of R_o ($v_{1,3}$) is very low. For $\psi < 1000$, the behaviour of the direction of maximal perceptual relevance is less consistent. The simulation results observed in the top panel of Fig. 2 correspond to experimental results in (Van Dinther et al., 2004). They found that parameter changes in the direction of the R_a parameter are most, R_k intermediate and R_o least important to create a perceptually discriminable vowel.

Van Dinther et al. (2004) also observed that the JNDs varied considerably between the three directions $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$. To measure the ratio of the perceptual effects of the three directions, the effects of R-parameter variations in the directions $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ are compared with the effects of variations in the most significant direction $\boldsymbol{v}_1$. In the middle panel of Fig. 2, these effects are quantified by the base-10 logarithms of the square-root eigenvalues ratios $\sqrt{\lambda_2/\lambda_1}$ (plus signs) and $\sqrt{\lambda_3/\lambda_1}$ (upward triangles). For, approximately, $\psi > 2000$, the square-root eigenvalue ratio $\sqrt{\lambda_2/\lambda_1}$ remains about constant. For $\psi > 3000$ the square-root eigenvalue ratio $\sqrt{\lambda_3/\lambda_1}$ remains about constant. This result indicates that for higher values of ψ the square-root eigenvalue ratios of the perceptual relevance of the three directions remain constant. For $\psi < 1000$ the square-root eigenvalue ratios and eigenvectors show more variation and jumps.

The method used for determining the perceptual relevance of the parameters can also be used to determine whether all three parameters are needed to model glottal pulses for small variations

of the R parameters. If one parameter direction is significantly more perceptually relevant than the other two, then this parameter could be sufficient to model small changes. The decision whether the LF model operates as a one-, two- or three-parameter model for these parameter vectors for small variations of the R parameters depends on a threshold for the square-root eigenvalues ratios. Veldhuis (1998) used a threshold value of 10% of the maximum value 1 (the maximum value 1 is reached when $\lambda_2 = \lambda_1$ or $\lambda_3 = \lambda_1$), to decide whether the LF model operates as an one or a two parameter model, when considering the spectral relevance of these parameters. When this threshold is used for the results in Fig. 2, it turns out that the LF model operates as a one parameter model for, approximately, $\psi > 1500$ and as a two- or three-parameter model for eigenvalues $\psi < 1500$. In any case, when the overall perceptual relevance ψ of the R parameter variations increases, only changes in one direction (or at most two) seem to remain perceptually relevant.

The bottom panel of Fig. 2 shows base-10 logarithms of the R parameters as a function of ψ. For the range 0–3000 the points of W_r are more or less equally spaced. A quarter of the points in W_r corresponds to a $\psi > 3000$. One can see in the panel that R_a, R_k and R_o (stars, squares and circles, respectively) show trends when parameterised by ψ and resemble points on a curve in $\mathscr{R}$. Considering the fact that the R-parameter vectors were taken from different sources comprising a variety of vectors obtained from distinct voice qualities, it is surprising that these *production* parameters can be ordered along a trajectory which is a function of a *perceptual* parameter.

3.2. The influence of F_0, level and formant frequencies

Since we observe this relation with ψ and the ordering of the voice-source parameters, it is not clear whether the ordering is influenced by different speech parameters. We therefore investigated how the order of the R-parameter vectors was influenced by other choices for the vocal tract filter, other values of F_0, whether the total spectral energy (TSE) was the same for all R-parameter vectors from the set W_r, and whether the TSE was kept constant during parameter variations. The amount of change of the ordering compared to the ordering of the default setting used in the former subsection, can be seen as an indication of the sensitivity for the particular speech parameter. The following parameters were considered which are presented in Table 2.

Figs. 3–6 show examples of results for different speech-parameter settings. In each of the figures one speech parameter was changed compared to the original configuration used in Fig. 2. In Fig. 3 the vowel was /i/, in Fig. 4 the F_0 was 220 Hz, in Fig. 5 the initial TSE of the speech signal across the elements in the set W_r was 55 dB sound pressure level (SPL), with a 0 dB reference of 20 μPa, and in Fig. 6 the TSE of the speech signal was held constant when the R parameters were varied. Figs. 3–6 have the same division into panels as Fig. 2.

The overall picture of the four figures reveals that the maximum value of ψ varies between 3300 and 7000. The top panels of the figures show that the direction of $\boldsymbol{v}_1$ remains constant for $\psi > 1000$ and is almost always in the direction of R_a. The top panel of Fig. 6 shows a different behaviour concerning the direction of maximal perceptual relevance. The direction of $\boldsymbol{v}_1$ is already constant for $\psi > 500$, but one has to consider that ψ has also the smallest range for this condition. Next, it can be seen in Fig. 6 that the contribution of parameter R_k in the direction of maximum perceptual relevance is less compared with the other figures. This agrees to experimental data of Henrich et al. (2003), where it was found that parameter variations in the direction of R_k are

Table 2
The speech parameters used for analysis

Varied speech parameters	Conditions
Vowel	/a/, /i/ or /u/
F_0	110 Hz or 220 Hz
Initial TSE across W_r	Variable or held constant
TSE during variations to R parameters	Variable or held constant

Parameters consist of vowels (3-valued), F_0 (2-valued), initial TSE across W_r (2-valued) and TSE of variations to the R parameters with respect to the TSE for a point in W_r (2-valued).

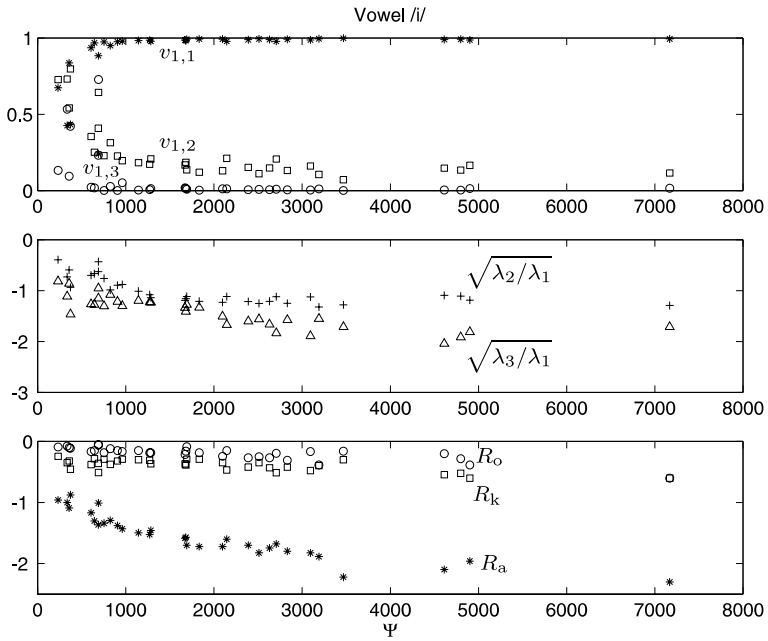

Fig. 3. One of four different speech-parameter realisations. The same construction is used as in Fig. 2. The following parameters are chosen for this figure: Vowel /i/, $F_0 = 110$ Hz, variable TSE across the set W_r and variable TSE to variations of R parameters.

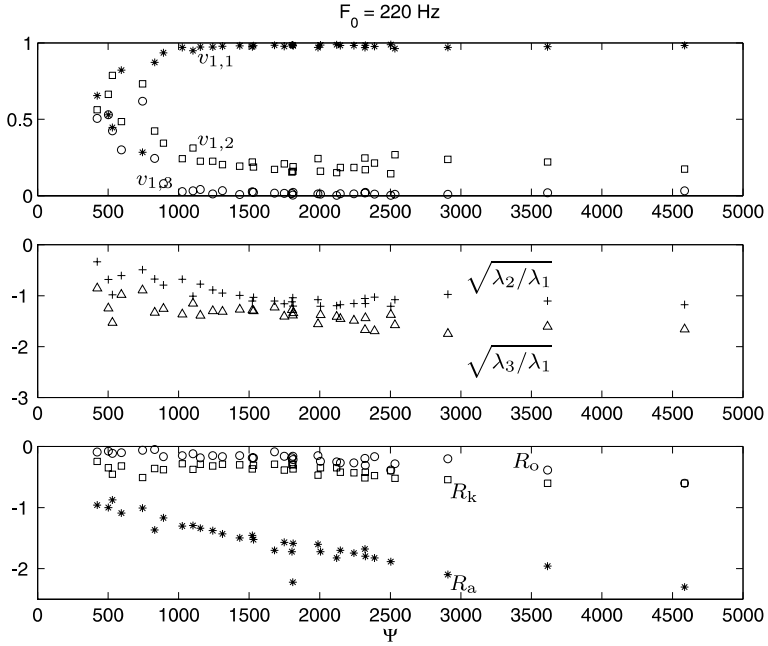

Fig. 4. One of four different speech-parameter realisations. The same construction is used as in Fig. 2. The following parameters are chosen for this figure: Vowel /a/, $F_0 = 220$ Hz, variable TSE across the set W_r and variable TSE to variations of R parameters.

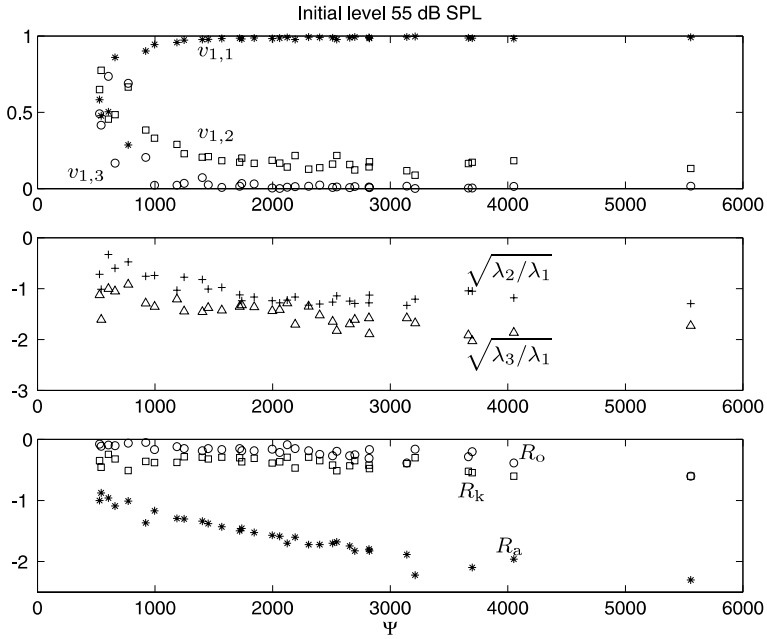

Fig. 5. One of four different speech-parameter realisations. The same construction is used as in Fig. 2. The following parameters are chosen for this figure: Vowel /a/, $F_0 = 110$ Hz, constant TSE across the set W_r and variable TSE to variations of R parameters.

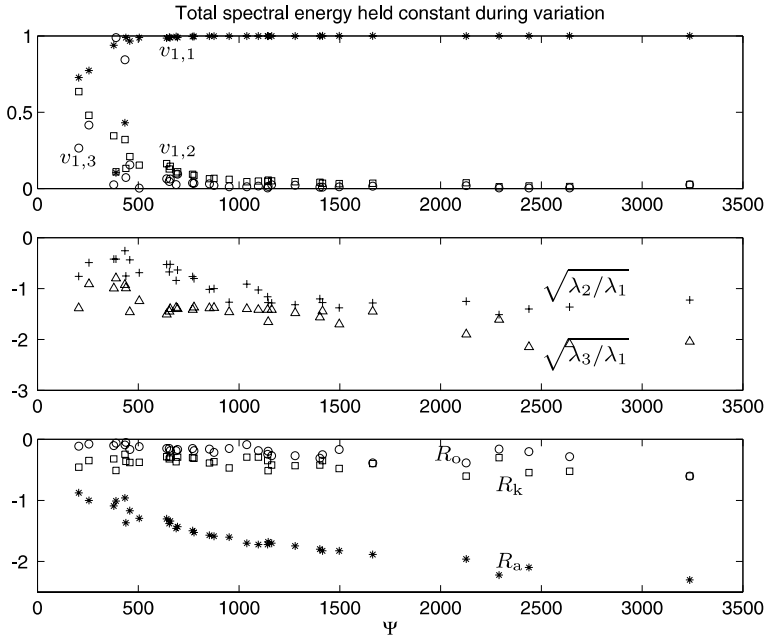

Fig. 6. One of four different speech-parameter realisations. The same construction is used as in Fig. 2. The following parameters are chosen for this figure: Vowel /a/, $F_0 = 110$ Hz, variable TSE across the set W_r and TSE held constant to variations of R parameters.

perceptually less relevant when the TSE is held constant. The middle panels of the figures show that for $\psi > 1500$ the square-root eigenvalue ratios stay below 10%, which means that the LF model operates as a one parameter model for these values of ψ.

The ordering of the R-parameter vectors along the perceptual parameter ψ presented in the bottom panels show some differences across conditions, but follows the same overall trend. To compare the different orderings for the different speech-parameter settings, we calculated the pairwise Spearman's rank correlation coefficients (Pestman, 1998) between these orderings of the elements of W_r. In total there are 24 different speech-parameter settings (i.e. 3 different vowels, 2 F_0 values, 2 different TSEs of the signals across R parameters and 2 different TSEs of the signals during variations, presented in Table 2), leading to $(24^2 - 24)/2 = 276$ pairwise correlation coefficients. The mean of the Spearman's rank correlation coefficients is high for the 276 pairs, namely 0.9641 with a standard deviation of the mean of 0.0017. In order to determine for which of the speech-parameter conditions presented in Table 2 the ordering is most invariant, the mean of the Spearman's rank correlation coefficients was calculated for those cases where only one speech parameter was varied. In Table 3 the means, the standard error of the mean and the medians are given for this analysis.

The means of Table 3 are all very high and above the grand mean (0.9641). Only the correlations between the conditions where the initial TSE of the signals of the corresponding R parameters was allowed to vary or was held constant at 55 dB SPL shows a clearly higher mean (0.9888), indicating that this speech parameter has the least influence on the ordering. The other three speech parameters have about the same mean.[1]

Table 3
The means, standard error of the means and the medians of the Spearman's rank correlation coefficients between orderings of R parameters when only one speech parameter is varied

Varied speech parameters	Mean	Standard error of the mean	Median
Vowel	0.9798	0.0042	0.9883
F_0	0.9695	0.0077	0.9796
Initial TSE across W_r	0.9888	0.0030	0.9925
TSE during variations to R parameters	0.9683	0.0079	0.9801

Resuming this subsection, it has been found that the orderings of the set W_r along ψ are pairwise highly correlated for different speech-parameter realisations and thus it can be assumed that the ordering of W_r along ψ is almost invariant to different speech-parameter realisations. The investigation is continued with the choice of speech parameters used before in Section 3.1, assuming that the findings of this particular case also hold for the other speech-parameter realisations. Because the elements W_r ordered along the parameter ψ are near a trajectory in $\mathscr{R}$, vectors of R parameters are investigated, which have been obtained from second-order approximations of the trajectories displayed in the bottom panel of Fig. 2. The results for these second-order approximations in relation with the parameter ψ will be discussed. This stylisation has also the advantage that R parameters can be obtained by a simple curve in $\mathscr{R}$.

4. Examination of the tracks of the stylised R parameters

In order to describe voice qualities by a single parameter it is desirable to have simple descriptions of the R parameters as function of this parameter. In this section a curve in $\mathscr{R}$ is introduced that can be used as a representation of the R parameters as a function of ψ.

In the bottom panel of Fig. 2 it was shown that the elements of the set of 33 parameters W_r ordered along the parameter ψ lie near a trajectory if the R parameters are considered in a logarithmic domain. It is assumed that the trajectories can be

[1] The Spearman's rank correlation coefficients have also been calculated for non-integer multiples of 110 Hz. The coefficients for the F_0 values of 200 Hz, 250 Hz and 300 Hz are 0.9792, 0.9567 and 0.9217, respectively. A comparison with the value for $F_0 = 220$ Hz (0.9695) indicates that the correlation value decreases systematically with increasing difference in the compared F_0 values.

sufficiently approximated by low-order polynomials. Therefore, second-order approximations are investigated in a least-squares sense, say $P^{a}(\psi)$, $P^{k}(\psi)$, $P^{o}(\psi)$, of the trajectories of the corresponding parameters R_a, R_k and R_o in this logarithmic domain. A curve γ for the representation of the R parameters is then composed by the second-order approximations as follows: $\gamma(\psi) := [10^{P^{a}(\psi)}, 10^{P^{k}(\psi)}, 10^{P^{o}(\psi)}]^{T}$. For an R-parameter vector on the curve, say $\gamma(\psi)$, the corresponding overall perceptual relevance, say ψ', can differ from the original ψ. A curve γ was found which is "invariant" to a certain degree, such that for a ψ and the corresponding overall perceptual relevance ψ', obtained from the parameter vector $\gamma(\psi)$, the perceptual distance between $\gamma(\psi)$ and $\gamma(\psi')$ is below the audibility threshold. The derivation of the curve is described in Appendix A.

In Fig. 7 the vectors ν_1 (top panel) and the base-10 logarithms of the square-root eigenvalue ratios (middle panel), computed from the R-parameter vectors on the curve ψ, are plotted. In the bottom panel, the R-parameter vectors from the curve γ ordered as function of ψ are plotted. The real R-parameter vectors of the set W_r with entries R_a, R_k and R_o are indicated with stars, squares and upper triangles.

The entries of the curve γ in the bottom panel and the entries of the eigenvector ν_1 in the top panel appear as smooth functions, which can be observed in the top and bottom panel of Fig. 7. Apart from the smoothness of the curves of Fig. 7, the results are similar to the results of Fig. 2. Again, from about $\psi > 1000$ the direction of maximum perceptual relevance remains constant, as can be observed in the top panel of the figure. The square-root eigenvalue ratios presented in the middle panel remain below 10% for $\psi > 1500$. The bottom panel of the figure shows that the R-parameter vectors on the curve γ are well ordered by the overall perceptual relevance. Furthermore, it can be seen that the entries of

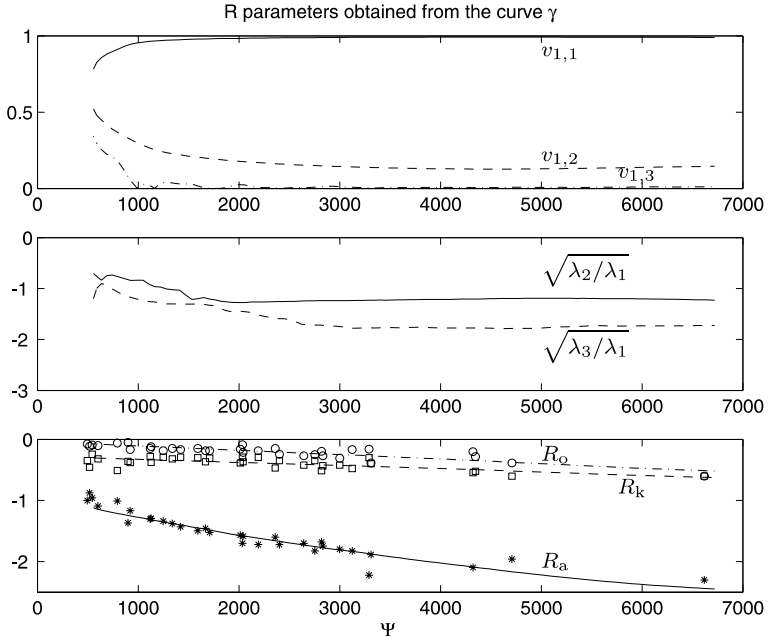

Fig. 7. Results for the vowel /a/, F_0 = 110 Hz for the R-parameter vectors on the curve γ. Top panel: Absolute values of the entries of ν_1 as functions of ψ; continuous, dashed and dashed-dotted lines correspond to the entries $v_{1,1}$, $v_{1,2}$ and $v_{1,3}$, respectively. Middle panel: Base-10 logarithms of $\sqrt{\lambda_2/\lambda_1}$ (upper line) and $\sqrt{\lambda_3/\lambda_1}$ (bottom line) as functions of ψ. Bottom panel: Base-10 logarithm of the R parameters as functions of ψ; R_o (dashed-dotted line), R_k (dashed line) and R_a (continuous line). The R parameters of the set W_r are also plotted; R_o (circles), R_k (squares) and R_a (stars).

the R-parameter vectors of the set W_r lie close to the stylised lines.

Until now, attention was paid to the contributions of the R parameters in the direction of maximum relevance $\boldsymbol{\nu}_1$. In Fig. 8 absolute values of the entries of all three eigenvectors $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$ and $\boldsymbol{\nu}_3$ are displayed for the curve $\boldsymbol{\gamma}$. The contributions of the eigenvectors of the set W_r are plotted as well in this figure; R_o (circles), R_k (squares) and R_a (stars).

It can be seen from the figure that the contributions of the eigenvectors calculated from the real parameter vectors follow the stylised lines closely. The middle and bottom panel shows that the significance of the contributions of the parameters R_k (dashed line) and R_o (dashed dotted line) shows a crossover near $\psi = 2000$. For $\psi > 2000$ the direction of $\boldsymbol{\nu}_2$ is mostly in the R_k direction and slightly in the other two parameter directions, while the direction of $\boldsymbol{\nu}_3$ is mostly in the R_o direction and slightly in the R_k direction. The contribution of R_a is low, but is more pronounced for $\psi < 2000$.

To find out the relations between the directions of maximum, intermediate and minimum relevance and the direction of the curve $\boldsymbol{\gamma}$, the inner products u_1, u_2 and u_3 of the orthonormal vectors $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$ and $\boldsymbol{\nu}_3$ have been calculated, respectively, with the tangent vectors $\dot{\boldsymbol{\gamma}}(\psi) := \boldsymbol{\gamma}'(\psi)/\|\boldsymbol{\gamma}'(\psi)\|$ of the curve $\boldsymbol{\gamma}$, where $\boldsymbol{\gamma}'(\psi) := [\gamma_1'(\psi), \gamma_2'(\psi), \gamma_3'(\psi)]^T$. The absolute values of the inner products are plotted in Fig. 9.

From this figure it can be observed that the direction of minimum relevance, $\boldsymbol{\nu}_3$ coincides with the vector $\dot{\boldsymbol{\gamma}}(\psi)$ for $\psi > 2000$, while the directions $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ coincide with a nearly orthogonal direction. For small ψ the direction $\boldsymbol{\nu}_2$ coincides with the vector $\dot{\boldsymbol{\gamma}}(\psi)$. The inner product with the vector $\boldsymbol{\nu}_1$ has a maximum at about the value $\psi = 1000$. The value of ψ where the lines of u_2 and u_3 intersect, corresponds to the value of ψ where the two lines of R_k and R_o in the middle and bottom panel of Fig. 8 intersect.

The image of the curve $\boldsymbol{\gamma}$ in $\mathscr{R}$ may be considered as a representative of all R parameters as

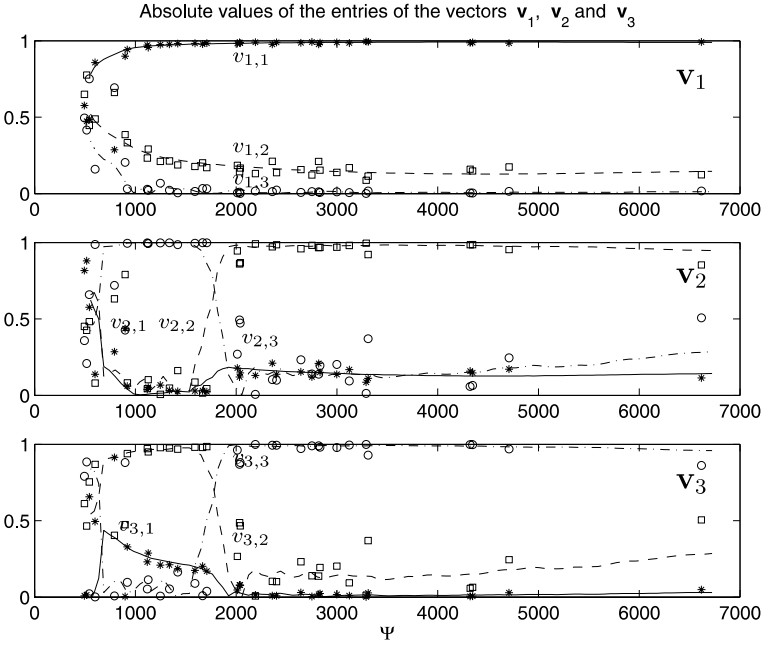

Fig. 8. Results for the vowel /a/, $F_0 = 110$ Hz for the R-parameter vectors on the curve $\boldsymbol{\gamma}$. Top panel: Absolute values of the entries of $\boldsymbol{\nu}_1$ as functions of ψ; Middle panel: Absolute values of the entries of ν_2 as functions of ψ; Bottom panel: Absolute values of the entries of $\boldsymbol{\nu}_3$ as functions of ψ; The continuous, dashed and dashed-dotted lines are the contributions of R_a, R_k and R_o, respectively. The contributions of the set W_r are also plotted; R_o (circles), R_k (squares) and R_a (stars).

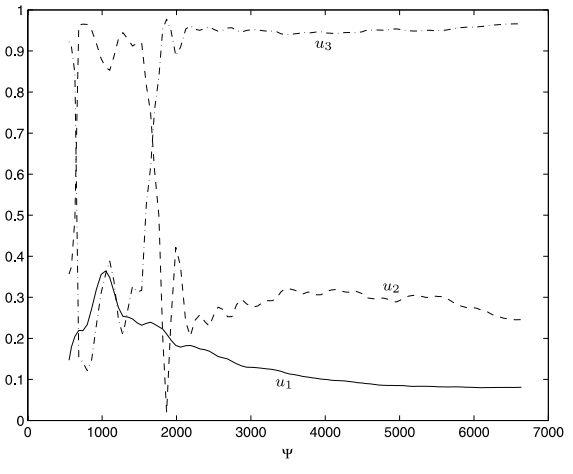

Fig. 9. Inner products u_1, u_2 and u_3 of the vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ with the tangent vectors of $\boldsymbol{\gamma}$, respectively.

function of ψ corresponding to various voice qualities. In Fig. 9 it is shown that for higher values of ψ (i.e. where the overall perceptual relevance is higher), the directions of maximum perceptual relevance are orthogonal to the directions of the curve. Further investigation is needed to find out whether the R-parameters changes in the orthogonal directions of the curve have an influence on voice quality.

In the next section the production parameter R_d is discussed, which can be used as a measure for describing voice qualities (Fant, 1993). It is shown that this parameter R_d is closely related to ψ.

5. Relation between the production parameter R_d and the perceptual parameter ψ

In the previous sections it was found that the R parameters can be parameterised by one perceptual parameter ψ. In Fant (1995) a statistical analysis of covariation of R parameters ranging from an adducted phonation to a breathy abducted phonation brought out characteristic trends which were quantified along a single shape parameter R_d, closely related to $T_d := U_0/E_e$, by $R_d := cF_0T_d = cF_0U_0/E_e$, where U_0 is the maximum airflow of the glottal pulse, E_e is the maximum excitation, and c a normalisation constant. Because R_d is a unique function of U_0, E_e and F_0, the parameter can be derived with simplified inverse filtering techniques (Fant, 1993). Additionally, R_d can also be calculated from the parameters R_a, R_k and R_o. Conversely, the R_d parameter permits the prediction of default R parameters from R_d (Fant, 1995). The predicted R parameters are denoted as $R_{ap} := (-1 + 4.8R_d)/100$; $R_{kp} := (22.4 + 11.8R_d)/100$; and $R_{op} := (1 + R_{kp})/2R_g$, where $R_g := R_{kp}/(4R_d/(0.5c + 1.2cR_{kp}) - 4R_{ap})$. It is interesting to compare this parameter R_d with the perceptual parameter ψ because of their seemingly equivalent prediction properties.

The values R_d and ψ have been computed for the R-parameter vectors on the curve $\boldsymbol{\gamma}$. These obtained values of R_d and ψ are plotted in Fig. 10, with R_d on the y-axis and ψ on the x-axis. The numbers in the plot are the R_d values and ψ computed from the 33 parameter vectors of W_r. It can be observed that there is a reciprocal relation between R_d and ψ. The range of R_d obtained from this analysis is about the main range of R_d described by Fant (1995). Values of R_d higher than this main range correspond to transitions into abducted termination of voicing. Because of the reciprocal relation between R_d and ψ, we can rephrase the conclusion which Fant (1995) has drawn on the relation between R_d and voice quality in terms of ψ, which is that voice quality at laryngeal level

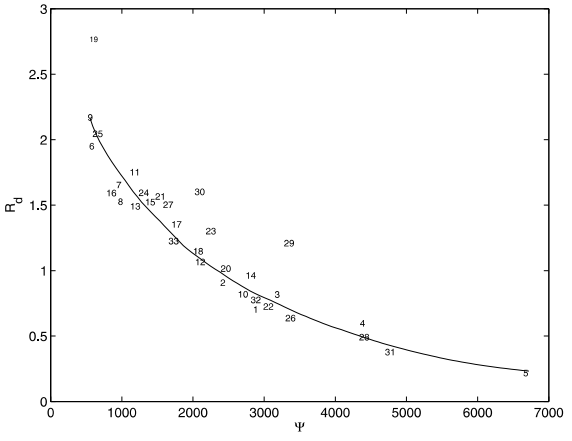

Fig. 10. Relation between the parameters R_d and ψ obtained from the R-parameters $\boldsymbol{\gamma}(\psi)$ (continuous line). The position of the numbers indicates the relation between R_d and ψ of the real R-parameter vectors from W_r.

can be parameterised by a single parameter. Furthermore, from the numbers in the figure corresponding to the voice qualities presented in Table 1 we can conclude that the categories of the various voice qualities are clustered. For values of ψ from 0 to 2000 the voice qualities breathy, lax and falsetto are mostly represented. For values of ψ between 2000 and 3000 the voice quality modal is represented. For values above 3000 the voice qualities tense and vocal fry are represented. Fant found that low R_d values are associated with low R_a and low R_o values. This corresponds to R-parameter settings of tense voice qualities. High values of R_d are associated with abducted phonation and high values of R_o, which corresponds to the voice qualities lax and breathy. Next, it can be observed that female voices generally have lower values of ψ than male voices. This may be due to the fact that female voices are more breathy than male voices. However, one must keep in mind that within gender there can be much more variation of breathiness and that some male voices can be more breathy than female voices (Klatt and Klatt, 1990). Also an individual is capable of producing a wide range of utterances differing in degree of breathiness. The low values of ψ for the R-parameter vectors corresponding to the voice quality falsetto are not that surprising, because the physiological characteristics of a gradual glottal opening and closing with short or no closed phase of the falsetto voice are similar to the voice qualities lax and breathiness which have incomplete glottal closure (Childers and Lee, 1991).

In order to investigate the perceptual relevance of small variations to predicted R parameters of the set W_r and the ordering of these predicted R parameters, the values R_d for the elements of W_r were calculated. Next, the set of predicted R parameters W_p were calculated from the values R_d corresponding to the set W_r. Finally, the ψ corresponding to the elements of the set W_p were calculated. In Fig. 11 the set W_p as function of ψ is shown. Again the same division into the panels is used as in Fig. 2.

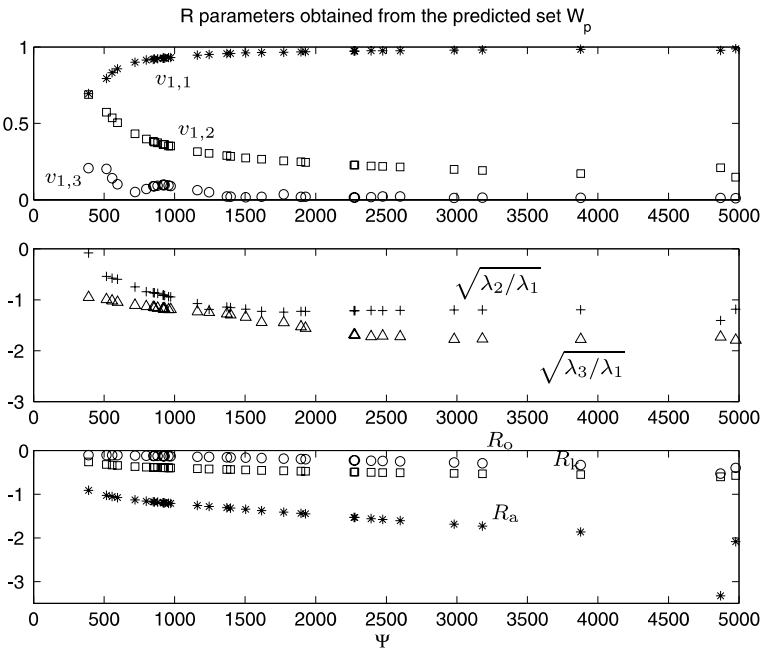

Fig. 11. Results for the vowel /a/, F_0 = 110 Hz for the R-parameter set W_p. Top panel: Absolute values of the entries of $\boldsymbol{v}_1$ as functions of ψ; the symbols star, square and circle, correspond to the entries $v_{1,1}$, $v_{1,2}$ and $v_{1,3}$, respectively. Middle panel: Base-10 logarithms of $\sqrt{\lambda_2/\lambda_1}$ (plus signs) and $\sqrt{\lambda_3/\lambda_1}$ (upper triangles) as functions of ψ. Bottom panel: Base-10 logarithm of the R parameters as functions of ψ; R_o (circles), R_k (squares) and R_a (stars).

The entries of the eigenvector ν_1 (top panel) and the logarithms of the elements of the set W_p (bottom panel) show an interesting result; the elements appear to lie near linear trajectories, except for one element in W_p. The R_{ap} of this element is a factor 10 smaller than the real parameter R_a, while the other predictions of the elements in W_r deviate only by factors between 0.5 and 1.5. This particular element in W_p corresponds to the prediction of an R-parameter vector corresponding to the voice type vocal fry (vowel /a/, $F_0 = 220$ Hz) obtained from (Childers and Lee, 1991). It is not clear why the prediction of this particular R-parameter vector shows a deviation.

In order to compare predictions from γ and from R_d with the set W_r of real R parameters, the entries of the elements in W_r are plotted against the entries of the elements in W_p and against $W_s := \{\gamma(\psi):\psi$ calculated for the elements in $W_r\}$. In the top-left panel of Fig. 12 the R_a parameters of the set W_r are plotted against W_p (circles) and W_s (triangles). In the top-right panel and bottom-left panel the parameters R_k and R_o are plotted for the three sets, respectively. A logarithmic scale is used on the axes.

In the three plots it can be observed that the circles and triangles are relatively close to the 45-degree line. The deviations from this line are about the same for both sets W_p and W_s. The perceptual differences between the elements of the sets W_r and W_p, W_r and W_s and W_p and W_s, however, are found to lie between 0 and 50 dB EPD, which means that the deviations can be audible. It has to be investigated whether these audible differences also result in different voice qualities.

According to Fant, the main importance of the R_d parameter is that it is the most effective single measure for quantifying voice qualities at laryngeal level. Second, the R_d simplifies the description of text-to-speech source rules. Again, it is interesting to see that the perceptual parameter ψ is closely related to the production parameter R_d. In addition, the perceptual parameter ψ adds a new dimension to the R-parameter vectors, because the parameter R_d adds a perceptual value to each of the production parameters. Both the parameters

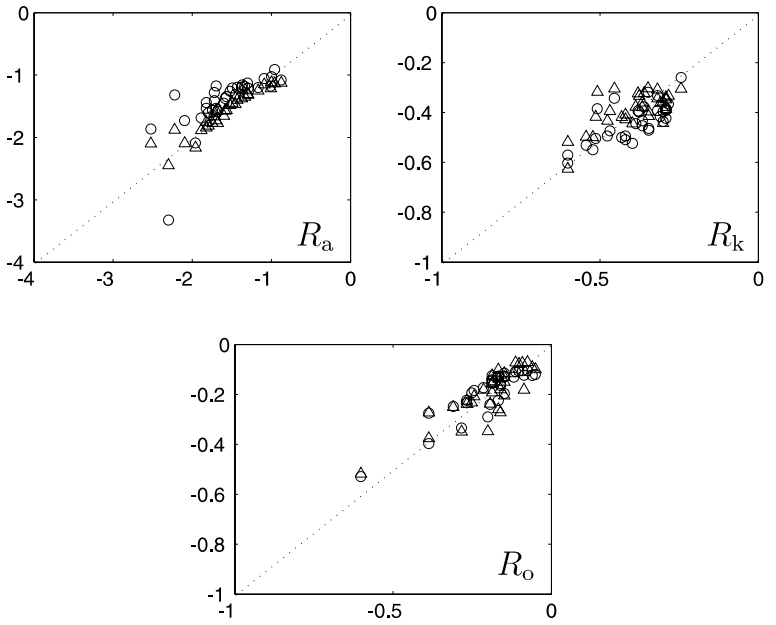

Fig. 12. Plots of the R-parameters R_a, R_k and R_o for the 33 parameter vectors (x-axis) against the elements of the predicted vectors W_p (circles) and $\gamma(\psi)$ (triangles), for values ψ calculated from the 33 parameter vectors. Top-left panel: Plots of the parameter R_a. Top-right panel: Plots of the parameter R_k. Bottom-left panel: Plots of the parameter R_o. On the x-axes and y-axes logarithmic scales are used.

R_d and ψ give a way to parameterise the three R parameters as function of a single parameter. The R_d parameter was statistically derived from the R parameters and gives no more information other than a correlation with the R parameters. The perceptual parameter, however, also adds information about auditory aspects. For example, the ψ parameter predicts the amount of variation of R parameters needed to reach a just noticeable difference. Consequently, it gives a way to divide the R parameter space into cells within which sets of R parameters are perceptually indistinguishable.

In Section 4 it was observed that the directions of highest perceptual relevance of the R parameters can be orthogonal to the curve γ. This raises the question whether such a parameterisation by the parameter R_d covers all the voice qualities. Next, the R-parameter vectors with higher values of ψ need more attention with respect to the classification of voice quality, where a small variation of the R parameters leads already to an audible difference.

6. Conclusions

This paper has revealed an unexpected relation between perception and production parameters of speech. It was found that the perceptual parameter ψ, which was derived from a perceptual distance measure applied to synthetic speech signals generated using different sets of production parameters, provides a link between speech production and perception.

It was demonstrated that the R parameters of the LF model can be approximated as simple functions of the perceptual parameter ψ. It is surprising that these production parameters can be ordered along a trajectory which is a function of this perceptual parameter.

For the higher values of ψ (i.e. $\psi > 1000$) the following observations were made. Firstly, the direction of maximum perceptual relevance becomes nearly constant and is mostly in the R_a and slightly in the R_k direction. Secondly, the LF model operates as a one- or two-parameter model, i.e. only changes in one direction (or at most two) seem to become perceptually relevant. In the case where the TSE of speech vowels were held constant for variations to R parameters the perceptual relevance in the direction of the parameter R_k decreases.

The perceptual effects were compared for several speech-parameter settings, such as different vowels and F_0. A Spearman's rank order correlation test showed high correlations between the speech parameters for orderings along the parameter ψ. This shows that the ordering of the R parameters along ψ is more or less invariant for speech-parameter variations.

The perceptual parameter ψ is approximately reciprocal with Fants basic shape parameter R_d. Because of the simple relation between ψ and R_d, the argumentation of Fant that R_d can be used as a measure for describing different voice qualities at laryngeal level holds also for the perceptual parameter ψ. In this paper it was concluded that the R-parameter vectors corresponding to different voice qualities are clustered when ordered as function of the perceptual parameter. For low values of ψ the voice qualities lax, breathy and falsetto are represented. For mediate values of ψ the voice quality modal is represented and for high values of ψ the voice qualities tense and vocal fry are represented. The question whether a curve γ in $\mathscr{R}$ covers all the voice qualities remains unanswered and deserves further investigation.

Acknowledgement

We would like to thank all reviewers for their valuable comments and remarks.

Appendix A. Derivation of the curve for representing the *R*-parameter vectors

In this appendix it is shown how the curve γ in $\mathscr{R}$ which was used in Section 4, is derived. Because the R-parameter vectors will be approximated in a logarithmic domain, iteration process is applied and analysed in a logarithmic domain obtained from a log mapping from the space $\mathscr{R}$. The set of R-parameter vectors in the logarithmic domain is defined as $\mathscr{R}_{\log} := \{[x, y, z]^T : 10^x, 10^y, 10^z \in \mathscr{R}\}$.

The set $\mathscr{R}_{\log}$ represents all the elements of $\mathscr{R}$ because it is assumed that $\mathscr{R} \subset \mathbb{R}^3 \setminus \{[x,y,z]^{\mathrm{T}} : x,y,z \leqslant 0\}$.

A curve in $\mathscr{R}_{\log}$ is defined by $\boldsymbol{\delta}(\psi) := [P^{\mathrm{a}}(\psi), P^{\mathrm{k}}(\psi), P^{\mathrm{o}}(\psi)]^{\mathrm{T}}$ as function of ψ, where $P^{\mathrm{a}}(\psi)$, $P^{\mathrm{k}}(\psi)$, $P^{\mathrm{o}}(\psi)$ are second-order approximations, in a least-squares sense, of the trajectories in $R_{\log}$ in the bottom panel of Fig. 2. For each vector in $\mathscr{R}_{\log}$ the corresponding value of the perceptual parameter ψ can be calculated, for convenience denoted by the mapping f: $\mathscr{R}_{\log} \rightarrow \mathbb{R}$. Because the curve $\boldsymbol{\delta}$ is an approximation of the trajectories, it is very likely that $f(\boldsymbol{\delta}(\psi)) \neq \psi$. If the elements $\boldsymbol{\delta}(\psi)$ are parameterised as function of the newly calculated $\psi' := f(\boldsymbol{\delta}(\psi))$, a different curve is obtained and the order of the elements $\boldsymbol{\delta}(\psi)$ can change. The effects considered above are investigated and a curve $\boldsymbol{\delta}_\infty$ is derived by iteration which is "invariant" with respect to the mapping f to a certain degree, i.e. such that the perceptual distance between $\boldsymbol{\delta}_\infty(\psi)$ and $\boldsymbol{\delta}_\infty(\psi')$ is well below the audibility threshold. We remark that it is desired that $\boldsymbol{\delta}_\infty$ and $\boldsymbol{\delta}$ are close in the sense that the perceptual distance between $\boldsymbol{\delta}_\infty(\psi)$ and $\boldsymbol{\delta}(\psi)$ for all ψ is below the audibility threshold, to maintain a good approximation of the trajectories presented in the bottom panel of Fig. 2.

The "second-order curve" (i.e. consisting of second-order polynomials) is derived by iteration. The "invariance" of a second-order curve $\boldsymbol{\delta}_n$ (the curve $\boldsymbol{\delta}_n$ after the nth iteration) is investigated at equally distributed mesh points $\{\psi_{0,0},\ldots,\psi_{K,0}\}$ with $\psi_{0,0} := \min(f(W_{\mathrm{r}}))$ and $\psi_{K,0} := \max(f(W_{\mathrm{r}}))$, where W_{r} is the set of the 33 real parameter vectors and $K = 100$. The step size between two consecutive mesh points $\psi_{i,0}$ and $\psi_{i+1,0}$ is chosen such that the perceptual distance between $\boldsymbol{\delta}(\psi_{i,0})$ and $\boldsymbol{\delta}(\psi_{i+1,0})$ is well below the audibility threshold. The iteration procedure is as follows. First, the tracks of Fig. 2 are approximated by the second-order curve $\boldsymbol{\delta}_1(\psi) := [P_1^{\mathrm{a}}(\psi), P_1^{\mathrm{k}}(\psi), P_1^{\mathrm{o}}(\psi)]^{\mathrm{T}}$. Second, parameter vectors $\boldsymbol{y}_{i,1} \in \mathscr{R}_{\log}$ are obtained by $y_{i,1} := \boldsymbol{\delta}_1(\psi_{i,0})$ at the mesh points $\psi_{i,0}$. The newly obtained parameter vectors $\boldsymbol{y}_{i,1}$ are parameterised by $\psi'_{i,1} := f(\boldsymbol{y}_{i,1})$. We can repeat this procedure by approximating this curve parameterised by the values $\psi'_{i,1}$, resulting in a curve $\boldsymbol{\delta}_2 := [P_2^{\mathrm{a}}(\psi), P_2^{\mathrm{k}}(\psi), P_2^{\mathrm{o}}(\psi)]^{\mathrm{T}}$, and so forth. By evaluating the curves $\boldsymbol{\delta}_1$, $\boldsymbol{\delta}_2$, ... at the mesh points $\psi_{i,0}$ after each iteration new parameter vectors $y_{i,1}, \boldsymbol{y}_{i,2}, \ldots$ are obtained, where $y_{i,n} := \boldsymbol{\delta}_n(\psi_{i,0})$, and corresponding values $\psi'_{i,1}, \psi'_{i,2}, \ldots$.

The parameter vectors $\boldsymbol{y}_{i,n}$ at the mesh points and the corresponding values $\psi'_{i,n} := f(\boldsymbol{\delta}_n(\psi_{i,0}))$ are calculated for $N = 100$ iterations. After each iteration the supremum norm $\|\ \|_\infty$ between vectors $\boldsymbol{\psi}_0 := [\psi_{0,0},\ldots,\psi_{K,0}]^{\mathrm{T}}$ and $\boldsymbol{\psi}'_n := [\psi'_{0,n},\ldots,\psi'_{K,n}]^{\mathrm{T}}$ is calculated. Next, the distances

$$\left(\int_{\psi_{0,0}}^{\psi_{K,0}} |\, P_{n+1}^{\mathrm{v}}(\psi) - P_n^{\mathrm{v}}(\psi)|^2 \,\mathrm{d}\psi\right)^{\frac{1}{2}} \tag{7}$$

are calculated for $n \in \{1,\ldots,N\}$ and $v = \mathrm{a},\mathrm{k},\mathrm{o}$.

The following results were observed after each iteration. First, it was found that the $\psi'_{i,n} > \psi'_{i+1,n}$ for all i and every n, which means that the order of $\psi'_{i,n}$ does not change. Consequently, parameter vectors obtained from the curve $\boldsymbol{\delta}_n$ stay in the same order once ordered again by the perceptual parameter ψ. Second, the perceptual distance between $\boldsymbol{\delta}_n(\psi_{i,0})$ and $\boldsymbol{\delta}_n(\psi_{i+1,0})$ is well below the audibility threshold for all i. Finally, the perceptual distance between $\boldsymbol{y}_{i,n+1}$ and $\boldsymbol{y}_{i,n}$ stays below the audibility threshold for $n \in \{0,\ldots,N-1\}$.

For iterations with $n \geqslant 50$ we have

$$\max_{v=a,i,u}\left\{\left(\int_{\psi_{0,0}}^{\psi_{K,0}} |\, P_{n+1}^{\mathrm{v}}(\psi) - P_n^{\mathrm{v}}(\psi)|^2 \,\mathrm{d}\psi\right)^{\frac{1}{2}}\right\} < 10^{-3}, \tag{8}$$

This means that the curves $\boldsymbol{\delta}_{n+1}$ and $\boldsymbol{\delta}_n$ are rather similar for large n. The sequence will probably not converge due to the condition that $\boldsymbol{\delta}_n$ is a second-order curve. Next, the supremum norm between the vectors $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}'_n$, $n \geqslant 50$ stays in the neighborhood of a constant. This constant value is low such that the difference between $\boldsymbol{\delta}_\infty(\psi)$ and $\boldsymbol{\delta}_\infty(\psi')$ is well below the audibility threshold at the mesh points $\{\psi_{0,0},\ldots,\psi_{K,0}\}$. The perceptual difference between $\boldsymbol{\delta}_{50}(\psi_{i,0})$ and $\boldsymbol{\delta}_{50}(\psi'_{i,50})$ is well below audibility threshold for all i. The perceptual distances between $\boldsymbol{\delta}_1(\psi_{i,0})$ and $\boldsymbol{\delta}_{50}(\psi_{i,0})$ are below threshold for almost every i, except for $i = 1,2$ and 3, where the threshold is about 1–1.5 dB EPD above the audibility threshold of 4.3 dB EPD.

The curve $\boldsymbol{\delta}_\infty := \boldsymbol{\delta}_{50}$ is used as a representation of the parameters in $\mathscr{R}_{\log}$ as function of ψ. A curve γ which represents the real R-parameter vectors can be simply obtained by taking the power with base 10 of the entries of $\boldsymbol{\delta}_\infty$.

References

Childers, D., Lee, C., 1991. Voice quality factors: analysis, synthesis and perception. J. Acoust. Soc. Am. 90, 2394–2410.

Fant, G., 1993. Some problems in voice source analysis. Speech Commun. 13, 7–22.

Fant, G., 1995. The LF-model revisited. Transformations and frequency domain analysis. STL-QPSR 2–3/95, 119–156.

Fant, G., Liljencrants, J., Lin, Q., 1985. A four-parameter model of glottal flow. STL-QPSR 4/85, 1–3.

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 17, 481–494.

Karlsson, I., Liljencrants, J., 1996. Diverse voice qualities: models and data. STL-QPSR 2/96, 143–146.

Kawahara, H., Masuda-Kasuse, I., de Cheveigne, A., 1999. Restructuring speech representations using pitch-adaptive time-frequency smoothing and instantaneous-frequency-based f_0 extraction: Possible role of repetitive structure in sounds. Speech Commun. 27, 187–204.

Kawahara, H., Matsui, H., 2003. Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. In: Proceedings of ICASSP'03, Vol. I. pp. 256–259.

Klatt, D., Klatt, L., 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. J. Acoust. Soc. Am. 87, 820–856.

Laver, J., 1980. The Phonetic Description of Voice Quality. Cambridge University Press, Cambridge.

McAulay, R., Quatieri, T., 1986. Speech analysis/synthesis based on a sinusoidal representation. IEEE Trans. Acoust. Speech Signal Process. 34, 744–754.

Moore, B., 1987. Distribution of auditory filter bandwidths at 2 kHz in young normal listeners. J. Acoust. Soc. Am. 81, 1633–1635.

Moore, B., 2003. An Introduction to the Psychology of Hearing, fifth ed. Academic Press, San Diego.

Moore, B., Glasberg, B., Baer, T., 1997. A model for the prediction of thresholds, loud-ness and partial loudness. J. Audio Eng. Soc. 45, 224–240.

Moulines, E., Charpentier, F., 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Commun. 9, 453–467.

Pestman, W., 1998. Mathematical Statistics. De Gruyter, Berlin.

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. Am. 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., 2004. A method for analysing the perceptual relevance of glottal-pulse parameter variations. Speech Commun. 42, 175–189.

Veldhuis, R., 1998. The spectral relevance of glottal-pulse parameters. In: Proceedings of ICASSP'98, Vol. II. pp. 873–876.