Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

LF-frequency domain analysis

Fant, G. and Gustafson, K.

journal: TMH-QPSR
volume: 37
number: 2
year: 1996
pages: 135-138

KTH Computer Science and Communication

http://www.speech.kth.se/qpsr

LF-frequency domain analysis

*Gunnar Fant and Kjell Gustafson**
*Dept. of Speech Music and Hearing, KTH. *Also Telia Promotor AB*

Abstract

An analysis-by-synthesis frequency domain procedure for deriving LF-voice source parameters has been developed. It is based on a detailed analysis of the frequency domain properties of the LF-model and synthesis requirements described in Fant (1995). It operates directly on harmonic spectra from an ordinary tape recording. An additional advantage is that the spectral matching is performed against a well defined synthesizer configuration which guarantees a correct resynthesis. Male and female source data are exemplified. The perceptual significance of parameter variations are discussed.

Introduction

Conventional inverse filtering requires a HiFi recording preserving a correct amplitude and phase response extending to very low frequencies. Since all perceptually important information is present in an ordinary sound recording it seems attractive to define voice qualities entirely from spectral data. Stevens & Hanson (1994) have accordingly introduced a set of spectral measures relating the amplitude of the voice fundamental H1 to that of the second harmonic H2 and to formant amplitudes and noting the first formant bandwidth. Our approach has been directed to a determination of LF parameters providing a best overall spectral match, including some of the Stevens & Hanson parameters as criteria.

Conventional methods have not been void of spectral criteria. In our routines the final determination of the Fa parameter is usually checked in the spectrum of the single voice period under analysis. Considerable differences may occur between time and frequency domain determinations of Fa. Other factors influencing the Fa measure are the number of higher formants cancelled in the inverse filtering and their positions. To ensure optimal resynthesis one can impose the rules for higher-pole correction of a particular synthesizer. In any case, because of the trading relationship between Fa and higher-pole corrections, it is important to specify the particular conditions.

The LF-model is quite flexible but cannot account for all details of a glottal flow pattern. In theory it should be possible to cancel intruding pole-zero pairs from nasalization or a subglottal coupling, but we lack systematic studies of their effects. Also, there are several nonlinear interaction effects which may influence the spectral slope and add spectral dips (Fant & Lin, 1988).

As a result we usually end up with ad hoc determinations of extra poles and zeros which can suit a resynthesis of a specific utterance but lie outside a general production theory. The LF-model suits some voices perfectly. In other instances we may want to optimize Fa to preserve a correct amplitude of a specific formant or formant region.

One limitation of the LF-model which we will discuss is that it does not allow for a distinct secondary excitation, sometimes to be seen at the instant of glottal opening. Another limitation is the need to study covariation of vocal tract transfer functions and source functions imposed by glottal articulation, e.g. in terms of first formant bandwidth and noise generation.

Spectral representations provide a natural reference for perceptual evaluations. We have performed some informal listening tests.

The transformed LF-model

The established parameters of the LF-model are Rk Rg and Ra describing the shape of a glottal pulse and Ee, the excitation amplitude as defined in Fig. 1. For a more detailed presentation, see Fant et al. (1985), Fant (1995). A low value of Rk=(Te-Tp)/Tp indicates a flow pulse skewed to the right, a high value of Rg=T0/2Tp a rapidly rising pulse and a high Ra=Ta/T0 a great relative length of the return phase.

The most important LF-parameter is Ra or rather Fa=F0/(2πRa) which describes the spectral tilt. Fa is the cut-off frequency of an equivalent first order low pass filter which provides an additional 6 dB/oct source slope.

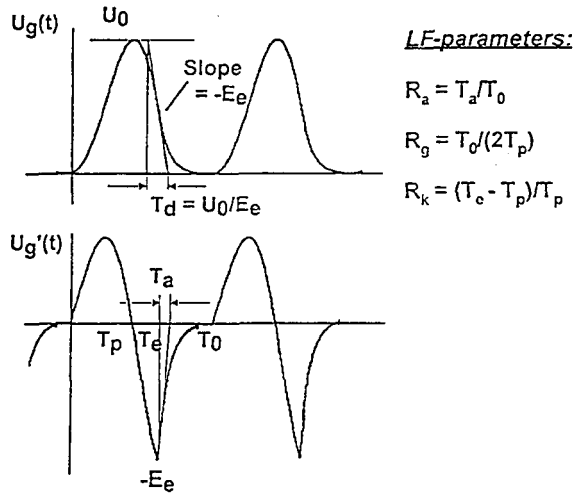

Figure 1. The LF voice source model.

The open quotient

OQ=(Te+Ta)/T0=(1+Rk)/2Rg+Ra (1)

is often defined so as to exclude Ra. This has been the practice in most of our publications and in the analysis of parametric interrelations. The new waveshape parameter Rd is defined as

Rd =(Uo/Ee)(F0/110)1000 (2)

if (Uo/Ee)=Td is expressed in seconds and as Rd=(Uo/Ee)/F0/110) with Td in ms. Alternatively, if the LF-parameters are known a good approximation to Rd is

Rd=(1/0.11)(0.5+1.2 Rk)(Rk/4Rg+Ra) (3)

The importance of the Rd-parameter is that it allows default predictions of Rk, Rg, and Ra labelled Rkp, Rgp and Rap. From statistical analysis we have found

Rap=(-1+4.8Rd)/100 (4)

Rkp=(22.4+11.8 Rd)/100 (5)

Rgp is obtained from Eq. 4 and 5 inserted into Eq 3.

Deviations from default values are expressed as

ka=Ra/(Rap) (6a)
kg=Rg/Rgp (6b)
kk=Rk/Rkp (6c)

where kk is a unique function of Rd, Ra and Rg and thus redundant

The shape vector [Rk, Rg, Ra] may thus be transformed to the more powerful vector [Rd, ka, kg], where the default values of ka and kg are equal to 1.

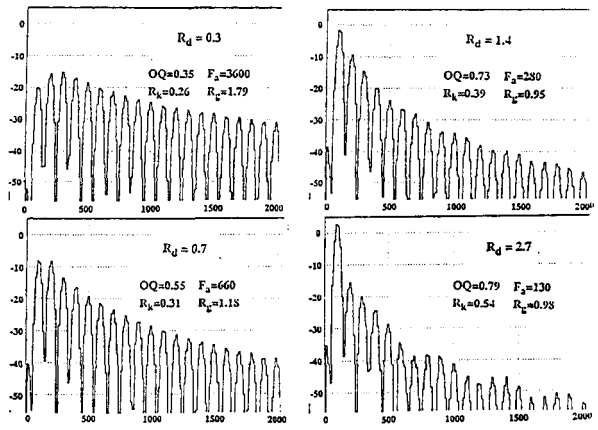

Figure 2. Source spectra at varying Rd.

Default source spectra for Rd=0.3, 0.7, 1.4, and 2.7, at F0=100 Hz are shown in Fig. 2.

The spectral correlates of the LF-parameters have been described in more detail in Fant (1955) than in earlier publications. It is thus shown that not only Rk and Rg but also Ra affect the lowest part of the spectrum at the voice fundamental and the lowest harmonics.

These relations provide a tie to the specificational system of Stevens & Hanson (1994). On a variational basis we may thus specify how great changes in each of Rk, Rg and Ra are needed to cause one decibel increase in the voice fundamental amplitude H1* and in H1*-H2*. The star indicates properties of the source spectrum, which can be recovered from the sound spectrum by a frequency domain undressing of the transfer function. The relations are summarized in the following table:

Table 1. Change in each of Ra, Rk and Rg needed to increase the level of the fundamental H1 by 1 dB and H1-H2 by 1 dB keeping other parameters constant.

Parameter	dR/dH1	dR/d(H1-H2)
Ra	1.0	1.25
Rk	3.0	4.5*
Rg	-12	-10

(*Observe a misprint in Fant, 1995)

Powerful analytical expressions also exist.

H1*-H2* = -6 + 0.27exp(5.5OQ) (7)

Here OQ is defined without Ra.
The linear relation

H1*-H2* = -7.6 + 11.1 Rd (8)

holds for moderate deviations from default parameters.

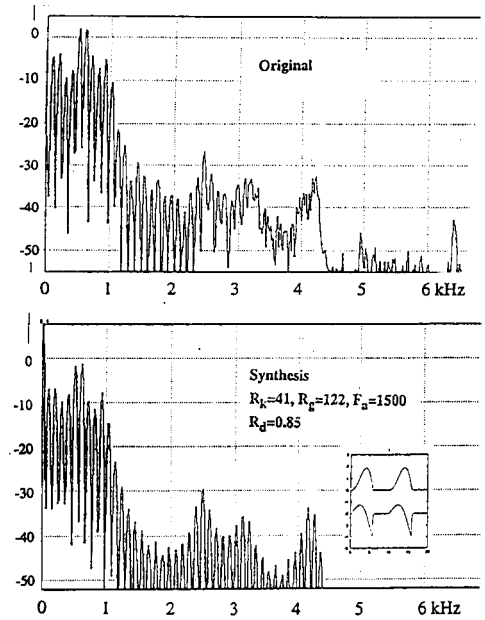

Figure 3. Spectral sections of a vowel [a] and a synthetic replica

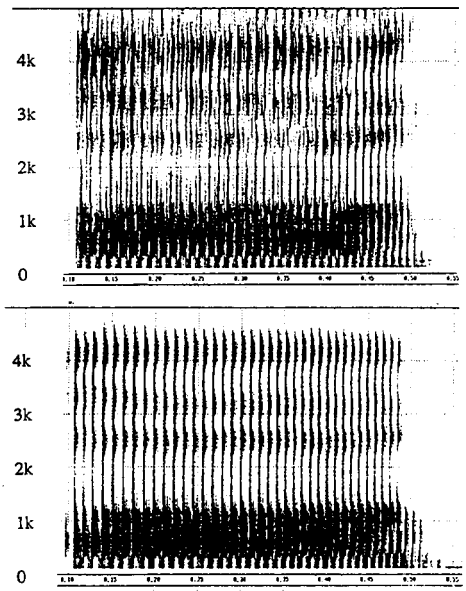

Figure 4. Spectrograms of natural and synthethic versions of the vowel [a]

Spectral matching

The analysis by synthesis is generally performed by matching of narrow-band spectral sections obtained by FFT over two successive voice periods. Initial estimates of formant frequencies and bandwidths can be supported by data from broad band spectrograms and automatic formant tracking. Initial estimates of LF-parameters are not crucial. Default values of Rk, Rg and Fa(Ra) corresponding to an expected Rd can be introduced. The F0 of the natural sample is transferred to the synthesizer and a first synthesis is carried out. Next, iterative corrections for the spectral difference between the natural and the synthetic sample are carried out by perturbing LF-parameters and formant frequencies and bandwidths.

Several variants of this strategy exist. The initial estimate of LF parameters may thus be based on the H1-H2 of the sound spectrum which by correction for the first and possibly also the second formant (see Eq. 11, page 127 of Fant, 1995), is converted to a corresponding measure H1*-H2* in the source spectrum from which Rd, Eq. (8) is solved followed by a calculation of the default values of Rk, Rg and Ra according to Eq. 3-5.

Alternatively, instead of resynthesis, the natural speech sample may be submitted to a regular inverse filtering preserving the synthesizer constraints. The spectral match is now performed in the source domain comparing spectral sections of the natural sample with reference data from a stored code book of source spectrum envelopes organized in terms of Rd, ka, and kg values. Fine adjustments can be made by reference to remaining errors in H1* and H1*-H2* converted to variations in LF-parameters according to Table 1.

Results from a spectral match of a vowel [ɑ] uttered by our reference subject ÅJ are shown in Fig. 3 and Fig. 4. The overall match between the natural sample and the GLOVE synthesis in the spectral sections of Fig. 3 is good up to F5 at 4200 Hz. The match gave Rg=122%, Rk=41%, Fa=1400Hz, Rd=0.86. With OQ´=(1+Rk)/2Rg= =0.58 inserted into Eq.7 we obtain H1*-H2*=0.6. Adding the contribution -1.2 dB of the transfer function, mainly the F1 influence, we predict H1-H2=-0.6 dB which is an exact match of the ÅJ sound spectrum.

Control determinations from conventional inverse filtering gave similar values but on the whole somewhat lower OQ, Rd, and H1*-H2*. These differences can be related to a rising zeroline in the maximally closed phase of the glottal flow which is ignored in the parameter extraction but causes a boosts in the voice fundamental

Female data

Successful frequency domain matching of female vowels up to F0=330 Hz have been attained.

Female voices show Rd values in the range of Rd=0.8-2.5 which overlaps the distribution Rd=0.5-1.5 typical of male vowels. Increasing Rd implies an increase of Rk and Ra, Fa decreasing and Rg on the whole decreasing.

Female voices usually have larger ka and thus lower Fa than men. This is especially true

of breathy, soft female voices, which also show a substantial glottal leakage and aspiration noise (Klatt et al., 1990, Karlsson, 1992).

Fine structure and perceptibility

A special study was devoted to the perceptibility of variations in the steady state LF-pattern. Informal listening of a systematically varied synthetic [ɑ] sound with constant Ee showed that there is a substantial tolerance for variations in Rk and Rg which primarily affect the low frequency region. Difference limen for H1* and H2* are of the order of 3 dB. The perceptually most important parameter is Fa in the range of Fa<1500 Hz and covarying variations in Rd>0.7. These findings confirm earlier evaluations in our department.

A detailed dynamic matching of source functions and formant patterns in about 16 frames covering the entire vowel of Fig. 4 was carried out. Correct onset and offset characteristics proved to be important for the perceived naturalness.

A specific feature often found in a detailed analysis is the presence of an extra excitation at the instant of glottal opening not predicted by the LF-model. This is to be seen in the spectrogram of Fig. 4. As a result there appears a fill in of the spectrum in the region of 1200-1800 Hz which apparently has a subglottal origin. It is also seen in the cross-sectional spectral view of Fig. 3. This distortion appears to be perceptually masked by the main formant structure.

The quasi-random fluctuations in the excitation of F3 and higher formants to bee seen in Fig. 4 probably add somewhat to the personal voice quality. This feature could partially be simulated by adding aspiration noise.

Acknowledgements

This work has been financed by grants from the Bank of Sweden Tercentenary Foundation, the Carl Trygger Foundation and support from Telia Promotor AB.

References

Fant G (1995). The LF-model revisited. Transformations and frequency domain analysis, *STL-QPSR* 2-3/1995: 119-156.

Fant G, Liljencrants J & Lin Q (1985). A four-parameter model of glottal flow, *STL-QPSR* 4/1985: 1-13.

Fant G & Lin Q (1988). Frequency domain interpretation and derivation of glottal flow parameters, *STL-QPSR* 2-3/1988: 1-21.

Karlsson I (1992). Modelling voice variations in female speech synthesis, *Speech Communication*, 11: 491-495.

Klatt D & Klatt L (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. *J Acoust Soc Am* 87: 820-857.

Stevens KN & Hanson M (1994). Classification of Glottal Vibration from Acoustic Measurements. In: Fujimura O & Hirano M, eds, *Vocal Fold Physiology 1994,* Singular Publ. Group. 147-170.

Ní Chasaide A, Gobl C & Monahan P (1994). Dynamic variation of the voice source in VCV sequences: intrinsic characteristics of selected vowels and consonants, *SPEECH MAPS (ESPRIT/BR No. 6975)* Delivery 15, Annex D.