EFFECTS OF VARYING LF PARAMETERS ON KLSYN88 SYNTHESIS

James Mahshie [1] and Christer Gobl [2]

[1] *Gallaudet University, Dept. of Audiology and Speech-Language Pathology, Washington, DC*

[2] *Centre for Language and Communication Studies, Trinity College, Dublin, Ireland*

ABSTRACT

The modified LF voice source model of KLSYN88 was investigated through spectral and glottal waveform analysis of synthesized vowels, where each of the four parameters of the model was systematically varied; open quotient (OQ), speed quotient (SQ), spectral tilt (TL), and the amplitude of voicing (AV). There was particular emphasis on comparison with the original LF model. The spectral analysis indicates that the effects of changing parameter settings are broadly in agreement with expectations. However, results for TL showed that attenuation is less than expected and there was an indirect effect on the excitation strength (EE). Pulse skew turned out to be always higher than expected from SQ, which was also found to vary with changes to other parameters. Furthermore, as the LF parameter EE is affected by the settings of not only AV, but also SQ and TL, the mapping from EE to AV is not straightforward.

1. INTRODUCTION

The Klatt synthesizer is a powerful, readily available synthesizer that has been used extensively by researchers and developers. A major improvement of the current version, KLSYN88 [1], compared to earlier version, KLSYN [2], is the provision of more elaborate voice source models. In addition to the impulse model incorporated in the earlier version, it contains two other models, namely the default model KLGLOTT88 and a modified version of the LF model [1, 3].

There is a growing body of descriptive research on voice source variation emerging, which uses the LF model as a way of describing and quantifying source characteristics (see [4] and references therein). Because KLSYN88 provides a means of systematically varying LF parameters, it provides a potentially useful means of testing the perceptual significance of data described in terms of these parameters.

A current obstacle to using KLSYN88 to study the perceptual effects of manipulating individual LF parameters lies in the differences between the original LF model and the modified implementation in KLSYN88. Variation of the LF parameters in KLSYN88 does not always yield the expected change in either the acoustic output or the differentiated glottal flow. It is not clear how measured LF parameters from speech analysis can be mapped to the source parameters in the KLSYN88 synthesizer to yield identical or equivalent waveforms. One example of this problem can be seen in Gobl and Ní Chasaide [5] where the glottal skew parameter in a perception test was probably not varied for the optimal range, as the mapping of the skew parameter was unexpected.

The objective, therefore, of the present study was to elucidate the mapping between the LF model and the implementation of the LF model in KLSYN88 (henceforth referred to as the Klatt-LF model). More specifically, this study explores the effects of manipulating the four control parameters of the Klatt-LF model (OQ, SQ, TL, and AV) to determine their effect on both the output spectrum and the voice source waveform. By comparing source waveform parameters for the original LF and the Klatt-LF models, the aim is to ultimately provide for direct mapping between the two. This should allow us to optimize the resynthesis of analytic data based on the LF model.

2. THE LF AND THE KLATT-LF MODELS

This section gives a brief summary of the main properties of the LF model, in both the original formulation and the Klatt implementation. For more detailed accounts see [1, 3, 6]. The LF model, illustrated in Figure 1, defines the differentiated glottal waveform in terms of two functions. During the open phase, which is the time from the glottal opening to the main excitation, the differentiated glottal waveform is determined by

$$U'_g(t) = E_0 \cdot e^{a t} \sin(2p \cdot \mathrm{FG} \cdot t) \,. \qquad (1)$$

This is essentially a 'sinusoid' whose amplitude grows exponentially. The return phase, i.e. the residual flow after the main excitation, is modelled by an exponential function. This function always yields zero amplitude at the point of the next glottal opening.

The function in formula (1) above corresponds to the output waveform of an underdamped resonator (negative bandwidth). In the Klatt-LF model the open phase is implemented using such a resonator, and the resonant frequency FG is determined directly from the parameters OQ and SQ (see definition below). Thus, the open phase can in principle yield identical waveforms in both implementations. The return phase, however, is implemented differently in the Klatt-LF model. The downward tilt of the source spectrum caused by the return phase is achieved by low-pass filtering the basic pulse shape using a critically damped resonator.

Another area where the two models differ concerns which aspects of the waveform are held constant and which are allowed to covary, as a particular parameter is altered. The LF model is typically described as a four parameter model, although it has five degrees of freedom. The remaining variable is set so that area balance is achieved given the values of the four specified parameters. This means that the positive part of the glottal pulse derivative has the same area as the negative part. In the Klatt-LF model, area balance is not considered. The parameters of the underdamped sinusoid are determined from SQ and OQ values using lookup tables.

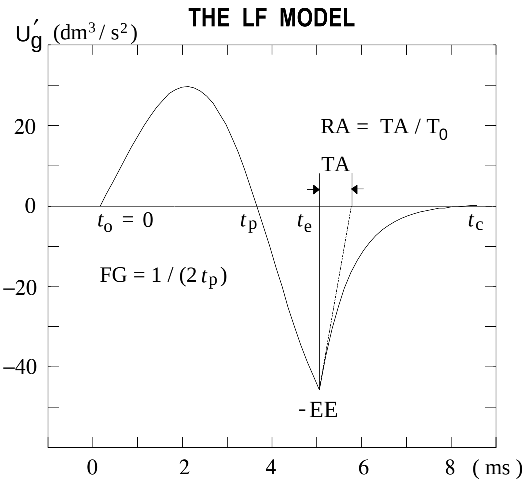

Figure 1. The LF model of differentiated glottal flow.

The LF parameters EE and RA are pertinent to the analysis that follows. EE, the excitation strength, corresponds to the amplitude of the glottal flow pulse at the point of maximum waveform discontinuity (see Figure 1). In the Klatt-LF model, EE is not independently controlled. Instead, amplitude is controlled by the parameter AV (amplitude of voicing), which is a simple scaling factor.

RA is the time constant of the return phase as a percentage of the glottal period, which determines the additional tilt of the source spectrum. In the Klatt-LF model, the cutoff frequency of the low-pass filter controlling the slope of the source spectrum is determined by the parameter TL (spectral tilt) on the basis of a lookup table. When TL > 10 dB, there is also an additional positive gain factor introduced in order to keep the amplitude at 300 Hz nearly constant. The numeric value of TL should correspond to the additional attenuation at 3 kHz.

SQ (speed quotient) is the skew of the pulse as determined by the ratio of the opening branch to the closing branch. OQ (open quotient) is the ratio of the open phase to the glottal period. SQ and OQ can in both models be directly determined by the time-points t_o (glottal opening), t_p (peak glottal flow), and t_e (time-point of excitation), see Figure 1.

To sum up, therefore, the Klatt-LF model is controlled in KLSYN88 by the parameters AV, TL, OQ, and SQ. Note that the latter two have precisely the same definition in the two models. The parameters AV and TL, however, although controlling similar aspects of the glottal pulse, do not match exactly their corresponding LF model parameters EE and RA.

3. PROCEDURES

3.1 Synthesis

A series of synthesized neutral vowels were created using the implementation of KLSYN88 contained in HLSYN (Sensimetrics Corporation, Boston, MA). The control parameters of the Klatt-LF model, AV, TL, SQ, and OQ were each independently varied along a range of values. This enabled examination of changes in the synthesized utterances resulting from manipulation of each parameter. A series of analyses were subsequently conducted on both the synthesized vowel waveforms and the voice source signals of the Klatt-LF model.

Table 1 shows the values used for TL, SQ, OQ, and AV. The default values, i.e. the values used for the parameters which were not varied in a particular series are underlined in Table 1. Other source parameters such as AH (aspiration noise), DI (diplophonia), FL (flutter) were not used and set to zero. Fundamental frequency was 100 Hz and formant values were set at the KLSYN88 default values, producing an /ə/ vowel quality. Five formants (in cascade) were used in the synthesis and the sampling rate was 11,025 Hz.

Parameter	Values
AV (dB)	30, 40, **50**, 60
TL (dB)	**0**, 5, 10, 15, 20, 30, 40
OQ (%)	30, 40, **50**, 65, 80, 95
SQ (%)	100, 175, **250**, 325, 450, 500

Table 1. Values varied independently in the synthetic vowels. Default settings underlined.

3.2. Spectral analysis

The first type of analysis examined the consequences of the various parameter manipulations on the vowel spectra. Examined were the amplitudes of the first harmonic (H1) and the first three formants (F1, F2, and F3). In the case where TL was varied, the amplitude of the harmonic at 3,000 Hz was also measured. These measures provided a means of examining the spectral changes associated with various Klatt-LF parameter manipulations to see if they are as would be expected on the basis of the known behavior of the LF model.

3.3. Glottal waveform analysis

The second type of analysis examined the glottal waveform characteristics of the Klatt-LF model for the parameter combinations synthesized. Using a series of programs that were developed at Trinity College, Dublin [7], the synthesized vowels were inverse filtered to obtain the source waveform, i.e. the Klatt-LF waveform. The LF model (in its original form) was then matched to the glottal pulses of the Klatt-LF model. The matching process involves manipulating the shape of the LF model by adjusting six cursors in the time domain, specifying five time-points and one amplitude point. From these points the LF model waveform is generated and superimposed on the Klatt-LF waveform. The matching software not only shows the match in the time domain, but also allows for comparisons of the corresponding spectra. The frequency domain matching is essential for achieving good estimates of certain source parameters, such as RA, and is particularly useful when trying to optimize the matching when a perfect temporal match is impossible.

4. RESULTS AND DISCUSSION

4.1 Spectral tilt, TL

Figure 2 shows the frequency and time domain consequences of altering the TL input parameter between 0 and 40 dB. As expected, the main effects of TL show up in the higher spectral regions. However, as can be seen in the top panel of Figure 2, the actual attenuation at 3 kHz does not match the TL value, but is rather less than would be expected. Thus, whereas the maximum TL input was 40 dB, a maximum of only 17 dB attenuation was observed at the harmonic nearest 3 kHz.

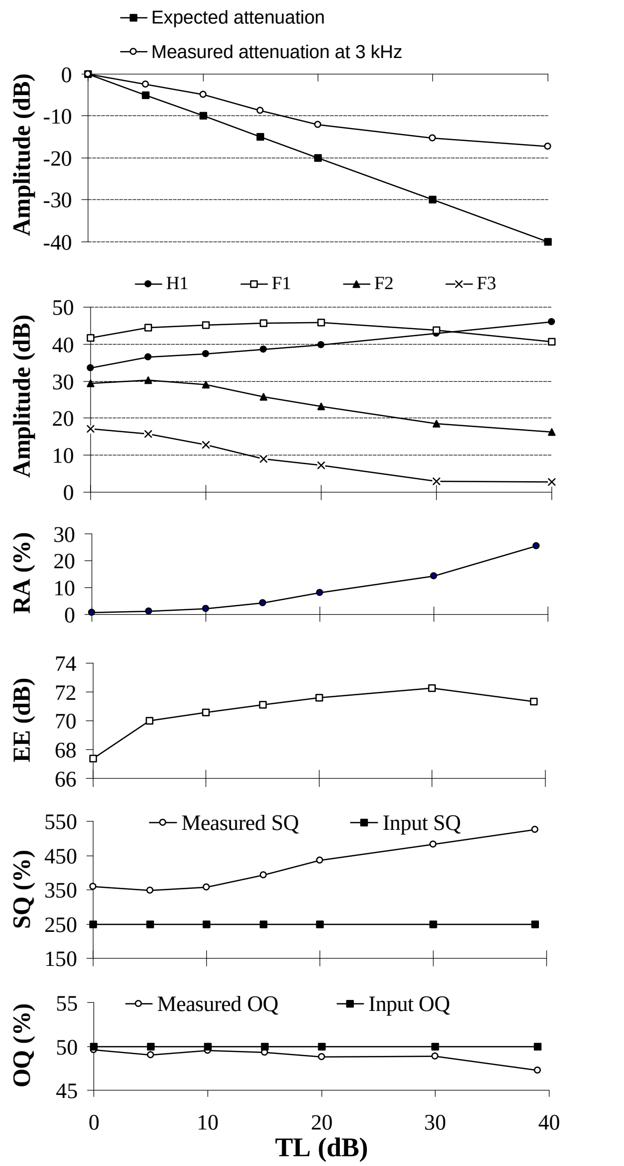

Figure 2. Variation in spectral amplitudes, RA, EE, SQ, and OQ as a function of the KLSYN88 parameter TL. The top panel compares actual additional attenuation to expected values.

Looking at the second panel of Figure 2, we see that for small values of TL (0 to 5 dB) only F3 is attenuated. The absence of an effect on F2 suggests that the cutoff frequency of the low-pass filter is higher than F_2 (1500 Hz). What is somewhat surprising in Figure 2 is the fact that H1, F1 and, to a slight degree F2 have a higher level for TL = 5 dB than for TL = 0 dB.

The reason for this becomes apparent when we look at what happens to EE when TL is varied. Except for very high TL values, EE rises with increasing TL values, and there is a striking rise of several dB when TL is changed from 0 to 5 dB. There is no obvious explanation for the sharp rise in EE when TL shifts from 0 to 5 dB. However, the rise in EE, for values above TL = 10 dB is probably due to the additional gain factor for TL > 10 dB. This may also help to explain why the level of H1 increases with higher TL values.

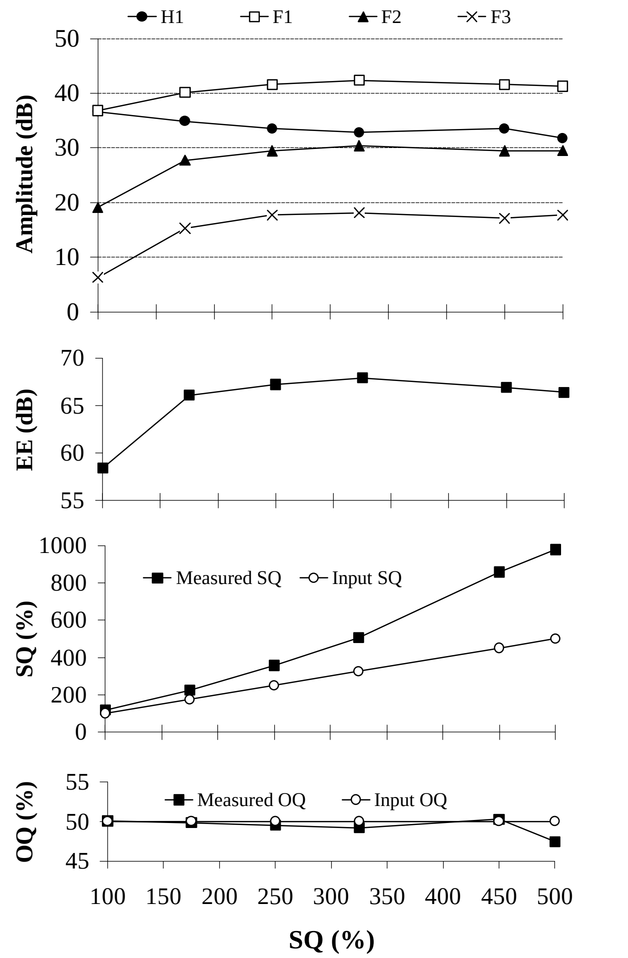

Figure 3. Variation in spectral amplitudes, EE, SQ, and OQ as a function of the KLSYN88 parameter SQ.

SQ is also affected by changes to TL. For an input SQ value of 250%, our measurements suggest that SQ is higher, and rises with increasing TL. For example, when TL = 10 dB, SQ is about 360% and when TL = 40 dB, SQ is about 525%. Input and measured OQ values turned out to be very similar, and OQ does not appear to vary much with TL variation. Taking SQ and OQ together, we can infer that the peak glottal flow occurs relatively later than would be expected.

Finally, the correspondence between TL and the equivalent LF parameter RA shows the expected trend. The amount of increase in RA with increasing TL is illustrated in Figure 2, and this plot should be useful for mapping between the two parameters.

4.2. Speed quotient, SQ

Figure 3 shows the effect on other source parameters and on spectral levels when SQ is varied between 100% and 500%. First of all, note again that SQ is higher than the specified input value, and this discrepancy increases with rising SQ. For an input SQ of 100% the output value was about 120%, but when the input value was 500% the output was close to 980%.

Increases at low SQ values (100% - 175%), result primarily in an increase of F2 and F3 amplitudes. A slight increase can also be seen for F1. Beyond SQ = 175% further changes in SQ has little effect on spectral levels.

The excitation strength, EE, varies as a function of SQ. Particularly at the lower end EE, rises sharply with increasing SQ. This sharp EE rise may explain the increase in the levels of F2, F3, and even F1. As before, the input and measured OQ values matched very closely.

4.3. Open quotient, OQ

Figure 4 illustrates the effect of OQ on spectral levels and on SQ, when OQ is varied from 30% to 95%. It is anticipated that changes in OQ would have a primary effect on frequencies below F1. For low values H1 amplitude is weak, and it rises as expected with increasing OQ. Changes to OQ did appear to cause relatively little alteration to other source parameters: EE was more or less constant, and input and measured OQ values corresponded well. The exception was SQ, which was, as always, higher than input values, and which dropped somewhat with increases in OQ.

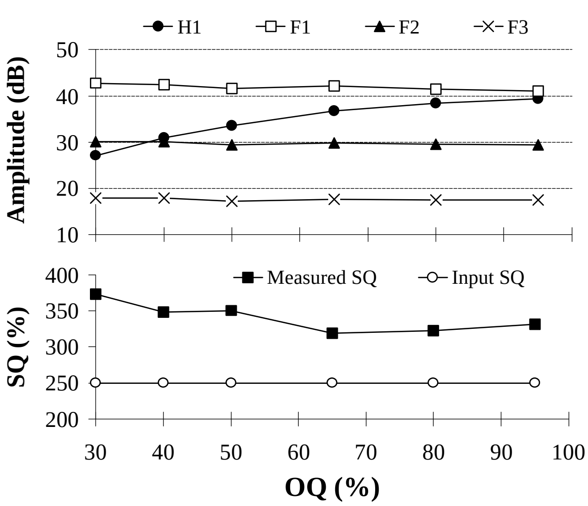

Figure 4. Variation in spectral amplitudes and SQ as a function of the KLSYN88 parameter OQ.

4.4 Amplitude of voicing, AV

The AV parameter behaved as expected, i.e. as a scaling factor. Increases in AV brought about a proportionate increase in the measured amplitude levels. In terms of the source parameters, a change in AV caused a corresponding change in EE, whereas the parameters RA, SQ, and OQ remained the same.

5. CONCLUSIONS

The modified LF voice source model of KLSYN88 (the Klatt-LF model) was investigated using spectral and glottal waveform analysis, with particular emphasis on comparison with the original LF model.

Results show that OQ behaves as expected and may generally be mapped directly from data based on the LF model. For the other Klatt-LF parameters TL, SQ, and AV, mapping is not as straightforward, and some surprising effects were discovered when TL and SQ were varied. Although the effect of increasing TL was indeed an increase in the spectral tilt, the level of additional attenuation was rather less than expected. From the correlation of TL and the corresponding LF parameter RA, it is clear that TL can be obtained from RA. However, it should be noted that the excitation strength, EE, also varies with TL. Measurements indicate that the pulse skew is always higher than expected on the basis of SQ values. The extent of the discrepancy also varies as a function of other parameter settings. As expected, AV works as a scaling factor. However, using EE data from the LF model to control AV is difficult and can give unpredicted results, as EE is affected by the settings of SQ and TL.

Further analyses are currently being carried out to try and establish robust procedures for transforming LF parameter data into appropriate settings for the Klatt-LF parameters.

ACKNOWLEDGMENTS

This work was supported through a Fulbright award to the first author. The research was conducted at Trinity College, Dublin and Gallaudet University in Washington.

REFERENCES

[1] Klatt, D.H. and Klatt, L.C. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-857.

[2] Klatt, D.H 1980. Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67, pp. 971-995.

[3] Fant, G., Liljencrants, J. and Lin, Q. 1985. A four-parameter model of glottal flow. *STL-QPSR*, Royal Institute of Technology, Stockholm, 4/1985, 1-13.

[4] Ní Chasaide, A. and Gobl, C. 1997. Voice source variation. *The Handbook of Phonetic Sciences*, eds. William J. Hardcastle and John Laver, Oxford, Blackwell, 427-461.

[5] Gobl, C. and Ní Chasaide 1999. Perceptual correlates of source parameters in breathy voice. *Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco.

[6] Fant, G. 1995. The LF-model revisited. Transformations and frequency domain analysis. *STL-QPSR*, Royal Institute of Technology, Stockholm, 2-3/1995, 119-156.

[7] Ní Chasaide, A., Gobl, C. and Monahan, P. 1992. A technique for analysing voice quality in pathological and normal speech. *Journal of Clinical Speech & Language Studies*, Dublin, 2, 1-16.