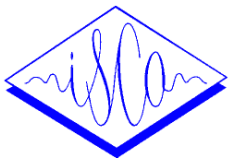

The ESCA Workshop
on Speech Synthesis
Autrans, France
September 25-28, 1990

HIGHER-LEVEL CONTROL PARAMETERS FOR A FORMANT SYNTHESIZER

Kenneth N. Stevens and Corine A. Bickley

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge MA 02139
and Sensimetrics Corporation, Cambridge MA 02139

ABSTRACT

Generation of speech with a formant synthesizer such as the Klatt KLSYN88 (Klatt and Klatt, 1990) requires that a large number of parameters (KL parameters) be manipulated. Many of these parameters are interrelated, and considerable precision may be required in specifying them. This paper describes a method of control of a formant synthesizer using higher-level (HL) parameters that are more independent of each other. These parameters are directly related to articulatory and aerodynamic variables, and have the advantage that they can be specified with less precision than the KL parameters.

1. INTRODUCTION

The control parameters for a formant synthesizer describe the movements of formants and pole-zero pairs, and the temporal characteristics of the glottal source and of turbulence noise sources. A formant synthesizer with flexibility to reproduce different voice qualities and with an adequate control inventory, such as the Klatt KLSYN88 synthesizer (Klatt and Klatt, 1990), requires more than 40 control parameters (called KL parameters here), and, consequently, many parameters must be specified by the rules for controlling the synthesizer from a phonetic sequence. Because of the constraints imposed by the articulatory system, the realistic combinations of these control parameters are highly restricted. In this paper we propose a set of 10 higher-level parameters (HL parameters) for controlling a formant synthesizer. This set is designed to simplify the rules for synthesizing utterances and to make the synthesis task more intuitive in terms of the relevant articulatory and acoustic processes involved in speech production. The 40-odd KL parameters are derived from these 10 HL parameters through a set of transformations.

The 10 HL parameters are: (1-4) The frequencies that the first four formants would have if there were no acoustic coupling to a side branch of the vocal tract; these natural frequencies are intended to be equivalent to a specification of the vocal-tract shape. (5-6) Parameters specifying the glottal frequency and degree of glottal abduction. (7) The degree of constriction of the vocal tract when that constriction is relatively narrow. (8) The cross-sectional area of the velopharyngeal opening. (9) The amount of active expansion of the vocal-tract volume, to control the intraoral pressure for obstruent consonants. (10) The degree of stridency, indicating the strength of turbulence noise at an obstacle in the vocal tract. In addition to these 10 higher-level parameters there would be several parameters that remain fixed over the duration of an utterance—parameters related to the individual characteristics of the speaker, the vocal effort, etc.

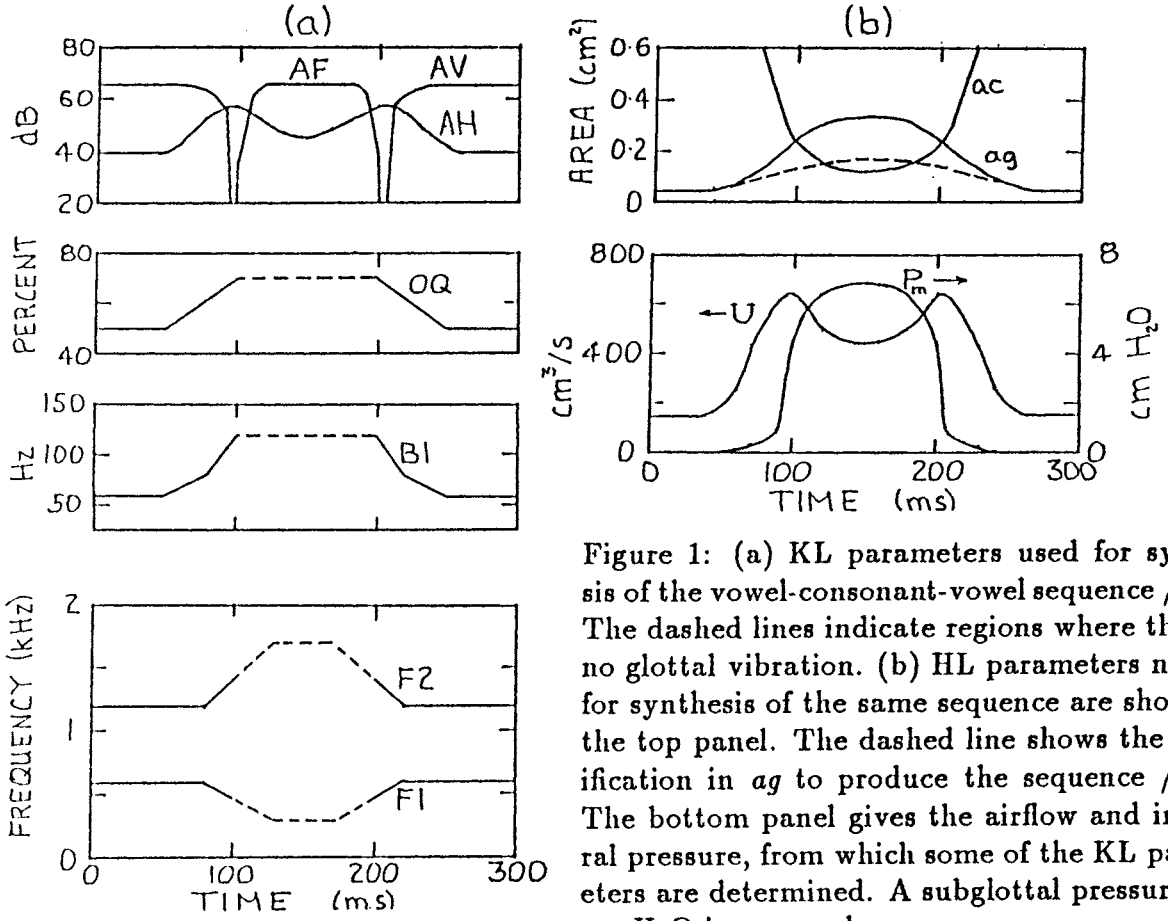

Figure 1: (a) KL parameters used for synthesis of the vowel-consonant-vowel sequence /ɑsɑ/. The dashed lines indicate regions where there is no glottal vibration. (b) HL parameters needed for synthesis of the same sequence are shown in the top panel. The dashed line shows the modification in *ag* to produce the sequence /ɑzɑ/. The bottom panel gives the airflow and intraoral pressure, from which some of the KL parameters are determined. A subglottal pressure of 8 cm H_2O is assumed.

2. EXAMPLES OF SYNTHESIS USING HL PARAMETERS

We illustrate the simplification provided by the use of HL parameters by comparing the array of KL and HL parameters that are needed to synthesize the utterance /ɑsɑ/. Figure 1a shows several of the KL parameters that are manipulated to produce this vowel-consonant-vowel utterance. A number of parameters have been omitted in order to simplify the figure; these include parameters to shape the spectrum of the noise and to modify the high-frequency tilt of the glottal spectrum.

Abrupt changes are evident in the KL parameters AV (amplitude of voicing) and AF (amplitude of frication noise), which turn on and off rapidly at the consonant boundaries. The peaks in AH (amplitude of aspiration) and the increase in OQ (open quotient) and in B1 (first-formant bandwidth) near the boundaries reflect the breathy voicing that occurs near the edges of the vowels adjacent to the fricative. These three changes, as well as the decrease in AV, result from a change in glottal configuration. All of these parameters, including the formant parameters at the bottom of the figure, must be carefully synchronized to produce the appropriate acoustic output, particularly the switching of the sources at the boundaries.

The time variations of the HL parameters that control the synthesis of this utterance /ɑsɑ/ are shown in Fig. 1b. We show in the top panel of the figure the two most relevant parameters, *ac* (cross-sectional area of supraglottal constriction) and *ag* (average area of glottal opening). In order to produce the intervocalic fricative, a supraglottal constriction is formed,

and, at the same time, the glottis is abducted. The glottal abduction is timed to begin before the supraglottal constriction is formed and extends beyond the release of the constriction. As can be shown, however, neither the precise value of *ag* during the frication interval nor the **timing** of *ag* in relation to *ac* significantly affects the synthesized speech.

The transformation or mapping of the HL parameters to the KL parameters that control individual components of the KLSYN88 synthesizer is based on known relations between vocal-tract shapes, aerodynamics, and acoustics (Fant, 1960). In the present example, the first step in this transformation is to calculate the intraoral pressure and the airflow through the glottis and through the supraglottal constriction (Rothenberg, 1968; Stevens, 1971), as shown in the second panel of Fig. 1b. The next step is to calculate the amplitude of the turbulence noise source at the supraglottal constriction knowing the airflow and the area, based on previous work on turbulence noise generation (Stevens, 1971; Shadle, 1985; Badin and Fant, 1969), to yield the KL parameter AF. The glottal source parameters are calculated next, and these are based on models of sound generation at the vocal folds. When the transglottal pressure decreases below a threshold value (about 3 cm H_2O), vocal-fold vibration ceases, and turbulence noise becomes the predominant source at the glottis. The results of these calculations are the KL parameters AV, AH, and OQ. The bandwidth parameter B1 also is determined from calculation of losses at the abducted glottis. This example illustrates the dependence of several KL parameters (AV, AH, OQ, and B1) on one articulatory event (spreading of the glottis). In contrast, the HL parameters are more independent of each other.

As a further example, we change the consonant from voiceless /s/ to voiced /z/ to produce the utterance /ɑzɑ/. If synthesis is based on the HL parameters in Fig. 1b, then just one change is necessary: *ag* is modified so that only a small amount of abduction occurs during the fricative, as shown by the dashed line in the figure. This single modification has several acoustic consequences that are taken care of by the HL–KL mapping relations: (1) the vocal folds continue to vibrate through all or part of the constricted interval; (2) the duration of interval of frication noise decreases; (3) the durations of the vowel portions of the utterances are slightly longer; and (4) there is little or no evidence of breathy voicing at the vowel boundaries. If the KL parameters were to be used to produce this utterance with a voiced fricative, a number of changes would need to be made in the parameters in Fig. 1a in order to achieve the proper acoustic output.

3. DISCUSSION

In summary, we have attempted to show the advantages in implementing a higher level of control for a formant synthesizer. Synthesis using HL parameters is more intuitive and simpler because: (1) less attention to interrelationships among parameters is needed, (2) a change in a single HL parameter produces a change in a single feature (such as voicing and manner), and (3) less precision in parameter specification is required.

The examples of synthesis of fricative consonants, together with other examples involving nasal and stop consonants (Stevens and Bickley, in press), have demonstrated that natural utterances can be generated using HL parameters, and have shown the advantages of this method of control. Some of the details of the HL-KL mapping relations have yet to be worked out and implemented, and the speaker-dependent aspects of these relations need to

be studied.

ACKNOWLEDGEMENT

This research was supported in part by the National Institutes of Health (Grant CD-00075 to the Massachusetts Institute of Technology and Grant NS-27407 to Sensimetrics Corporation).

REFERENCES

Badin, P. and Fant, G. (1989), "Fricative production modelling: Aerodynamic and acoustic data." In Proceedings, European Conference on Speech Communication and Technology, Eurospeech 89, Paris, 2, pp. 23-26.

Fant, G. (1960), Acoustic theory of speech production. The Hague: Mouton.

Klatt, D.H. (1980), "Software for a cascade/parallel formant synthesizer," Journal of the Acoustical Society of America, 67, 971-995.

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.

Rothenberg, M. (1968), "The breath-stream dynamics of simple-released-plosive production," Bibliotheca Phonetica, 6. Basel: S. Karger.

Shadle, C. (1985), "The acoustics of fricative consonants," Research Laboratory of Electronics Technical Report, 506, Massachusetts Institute of Technology, Cambridge MA.

Stevens, K.N. (1971), "Airflow and turbulence noise for fricative and stop consonants," Journal of the Acoustical Society of America, 50, 1180-1192.

Stevens, K.N. and Bickley, C.A. (in press), "Constraints among parameters simplify control of Klatt formant synthesizer," J. Phonetics.