ISCA Archive
http://www.isca-speech.org/archive

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

INVENTORY OF PHONETIC CONTRASTS GENERATED BY HIGH-LEVEL CONTROL OF A FORMANT SYNTHESIZER

David R. Williams, Corine A. Bickley and Kenneth N. Stevens

Sensimetrics Corporation
64 Sidney Street, Cambridge, MA 01821 USA

10.21437/ICSLP.1992-132

ABSTRACT

A method for simplifying the control of a Klatt-type synthesizer is described in which values of a small set of high-level (HL) parameters are transformed into values for the larger set of low-level (LL) acoustic parameters that typify formant synthesizers. The HL parameters include natural frequencies of the vocal tract, area of oral constriction, specification of a nasal side branch, active control of intraoral pressure for obstruents, and state of the glottal source. Determination of the LL source parameters involves the calculation of a set of intermediate parameter values which reflect the pressures and flows within the vocal tract. To illustrate the method, we describe the synthesis of speechlike sounds in which the HL orifice parameters are varied through their ranges of values, as well as the synthesis of a number of consonants in medial (VCV) position.

I. INTRODUCTION

The synthesis of speech with a terminal analog, or formant, synthesizer is based on a source-filter model of speech production [1]. Speech is synthesized by controlling parameters that specify various source characteristics, on the one hand, and the characteristics of time-varying filters that shape the sources, on the other [2][3]. In the case of nonnasal vowels, an all-pole filter is a good approximation to the vocal-tract transfer function, and such a filter is at the core of any formant synthesizer. Synthesis of other classes of sounds requires that the transfer function be expanded to have zeroes as well as poles, or that only the resonances of a portion of the vocal tract be specified.

Since the original conception of the formant synthesizer in the 1950's [4], advances in our knowledge of acoustic phonetics and in our ability to implement digital forms of the synthesizer have led to a number of improvements in formant synthesis. As a consequence of these advances, it is now possible to synthesize an utterance produced by almost any talker, male or female, so that the synthesized utterance is essentially indistinguishable from the original. These advances are not without cost, however. A modern formant synthesizer may have 40-odd parameters that can be manipulated, and synthesizing an utterance "by hand" is not only tedious but requires a considerable amount of detailed acoustic-phonetic knowledge on the part of the user.

II. THE HL APPROACH TO SPEECH SYNTHESIS

2.1 Higher-level control of formant synthesizer

One motivation for seeking an alternative approach to speech synthesis is to make it easier to control formant synthesizers. In the HL approach to speech synthesis, the values of a small set of high-level parameters are mapped into values for the larger set of low-level acoustic parameters which typify formant synthesizers like KLSYN88 [3]. Fewer HL parameters can be used because the relationship of these parameters to the synthesizer differs from that of the LL parameters: whereas the LL parameters determine the characteristics of the synthesizer sources and filters, the HL parameters determine values for *sets* of related LL parameters. This higher level of synthesizer control is established by a set of mapping equations. As our implementation of this extension to KLSYN88 is now essentially complete, we have been exploring and refining its capabilities by synthesizing a number of simple utterances [5].

The HL parameters include: (1) the vocal-tract shape, as defined by the natural resonances (*f1, f2, f3, f4*) of the vocal tract if there were no acoustic coupling to the trachea or to the nasal cavity; (2) the state of the glottal source as described by the fundamental frequency (*f0*) and by the average area (*ag*) of the glottal opening; (3) the area of the oral constriction (*ac*) when this constriction is narrow; (4) specification of the cross-sectional area (*an*) of the opening to the nasal side branch; (5) active control of intraoral pressure for obstruents, defined as the active expansion or contraction (*ue*) of the vocal-tract volume during obstruent consonants; and (6) a means for modifying the efficiency of turbulence noise generation for fricatives (*st*).

Values for the LL source parameters are determined by the pressures across orifices in the vocal tract and by the sizes of these orifices, as specified primarily by *ag* and *ac*. An initial step in determining the LL source parameters when the HL parameters are specified is to calculate the pressures across the orifices and the flows through them. A simple, low-frequency equivalent circuit model of the aerodynamics of the vocal tract forms the basis for these calculations [6]. From these intermediate parameters, it is then possible to calculate the amplitudes of the noise sources, and the characteristics of the periodic glottal waveform if the aerodynamic conditions permit glottal vibration.

The LL parameters which describe the filtering of the sources are determined from the vocal-tract shape (specified indirectly by the natural frequencies *f1, f2, f3,* and *f4*) and by the source locations and orifice sizes. For oral vowels with a modal glottal source, the natural fequencies are identical to the formant frequencies in the cascade branch of the formant synthesizer. When there is a wide glottal opening *ag* or a nasal opening *an*, the mapping equations modify the formant frequencies and bandwidths, based on acoustic theory that is reasonably well established [1][7]. A pole-zero pair is also introduced into the cascade branch when *an* is nonzero [8]. When there is a supraglottal noise source, filtering is achieved by setting the gains for a parallel array of formant resonators. The setting of these gains is dependent on the pattern of formant frequencies at the time the frication noise source is active.

2.2 Use of HL parameters in synthesis

While the use of HL parameters in speech synthesis has certain advantages over conventional approaches to synthesis, the actual process of formulating a set of parameters for synthesizing an utterance is somewhat less direct. In the conventional case, values for the synthesis parameters can, for the most part, be determined directly from acoustic analysis of naturally-produced utterances. For HL synthesis, some of the parameters, such as *ag, an, ac* and *ue*, are more directly related to articulation, and the time course of these parameters must be inferred from a knowledge of relations between articulation, airflow, and acoustics.

In order to determine how these articulatory-related parameters are to be manipulated in the synthesizer, one is guided by several kinds of information that are available in the literature. For example, there are several reports of direct observations of articulatory and glottal movements during consonant production. Of particular relevance for stop consonant production are the data of Fujimura, Kent and Moll, Perkell and others for articulatory movements [9][10][11], as well as various studies of laryngeal movements using fiberoptic or transillumination techniques. When there are narrow constrictions in the vocal tract, cineradiographic techniques are not sufficiently accurate, and electropalatographic methods can be used to estimate the kinematics of constriction configurations. For obstruent consonants, however, measurements of airflows and pressures probably provide a more accurate way of estimating constriction sizes than direct articulatory observations [12][13].

While these kinds of articulatory and aerodynamic measurements provide a guide to the estimation of certain HL parameters, the final criteria for fine-tuning the parameters are the properties of the sound that results from manipulation of the parameters. The time variation of the various constriction sizes must be adjusted so that acoustic attributes relating to frication noise, aspiration noise, glottal waveform, formant bandwidths, and nasalization are adequately matched to similar attributes observed in natural speech.

III. EXAMPLES OF HL SYNTHESIS

To illustrate the HL approach to synthesis, we consider in this section a series of synthesis examples. The examples demonstrate the acoustic consequences of manipulating the HL orifice parameters (*ag, ac, an*) and the natural frequency parameters, and describe how articulatory mechanisms involved in the production of various classes of English consonants can be implemented using HL parameters. The discussion also illustrates the methodology we are following to create an inventory of HL synthesis examples based on phonetic contrasts.

3.1 HL parameter *ag*

The spectrogram in Figure 1 shows the acoustic results of varying the parameter *ag* between values of 17 mm^2 and 0 mm^2 in unit steps with the parameter values held constant at each value for 40 msec. For this example, *ac* is set at 100 mm^2 (specifying an open vocal-tract configuration), the natural frequency parameters are set at values appropriate for the vowel [a] (*f1* = 700 Hz,

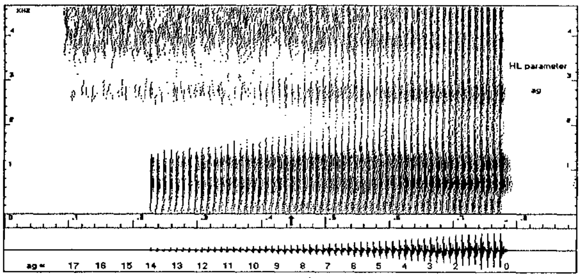

Fig. 1 - Spectrogram of sound produced by the synthesizer when the HL parameter ag *(glottal area) is decreased in steps of 1 mm^2 from 17 mm^2 to 0 mm^2. See text.*

f2 = 1100 Hz, *f3* = 2500 Hz, *f4* = 3500 Hz), and there is no coupling to the nasal side branch. The decrease in *ag* simulates a decrease in the glottal width from a relatively open position to a completely adducted position. Across this range of *ag* values, the HL mapping equations generate LL specifications that span a number of voicing types, including unvoiced aspiration (17 to 15), breathy voicing (14 to 6), modal voicing (5 to 3), pressed voicing (2 to 1), and finally, silence (0).

3.2 Glottal stop and aspirates

We next examine how the acoustic attributes associated with changes in the parameter *ag* are incorporated in synthesis. Figure 2 displays values of *ag* that we have used to synthesize a steady-state vowel [a] and three glottal consonants in intervocalic position: a glottal stop [ʔ] and voiced [ɦ] and voiceless [h] glottal aspirates. For each of these synthesized utterances, the shape of the supraglottal vocal tract is assumed to remain in a fixed position, as reflected by constant values of the natural frequency parameters for the vowel [a] (not shown in the figure).

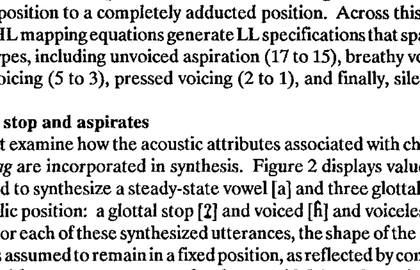

Fig. 2 - *Time course of the HL parameter* ag *(area of the glottal opening) used to synthesize the utterances [a], [aʔa], [aɦa], and [aha].*

In the synthesis of an intervocalic glottal stop [ʔ], we cause *ag* to decrease to zero at time t_1 and increase again to the modal width at t_2. The time interval within which *ag* moves from a value of 4 mm^2 to zero has been taken to be 30 ms in our synthesis of intervocalic glottal stops; the same time is assumed for the glottal opening movement. During these intervals, the LL parameters AV (amplitude of glottal vibration), OQ (open quotient) and TL (spectral tilt) are modified (through the HL mapping equations) to reflect the transition from modal through pressed voicing to termination of voicing, and the reverse.

To produce the glottal aspirates, the vocal folds undergo partial or complete abduction (for [ɦ] or [h], respectively). This movement is represented in HL synthesis by an increase in *ag*. For the voiced glottal aspirate [ɦ], *ag* increases from the modal width (*ag* = 4 mm^2) to a width of about 10 mm^2. At this width, the glottal source continues to vibrate, but there is a substantial increase in the calculated air flow. The acoustic result is voice-modulated noise. The HL mapping equations give rise to adjustments in the parameters OQ, TL, and AH (amplitude of aspiration noise). An increased first formant frequency and widened first formant bandwidth (due to the open glottis) are also introduced. As the glottal opening increases beyond an area of about 14 mm^2, glottal vibration is no longer sustained, and modulation of the noise ceases. For the voiceless glottal aspirate [h], *ag* may be as large as 30 to 40 mm^2. At this width, the mapping relations lead to estimates of a significant amount of turbulence noise at the glottis by adjustment of the LL parameter AH.

3.3 HL parameter *ac*

The spectrogram in Figure 3 shows the acoustic results of varying the parameter *ac* between values of 0 and 40 mm^2 in steps of 2 mm^2 with the parameter values held constant at each step for 20 ms. For this example, *ag* is set at a constant value (4 mm^2) and the nasal sidebranch is closed. At onset, the natural frequencies specify an alveolar closure; the frequencies of *f2* and *f3* move toward a high front vowel at offset. For most of the example (until *ac* reaches

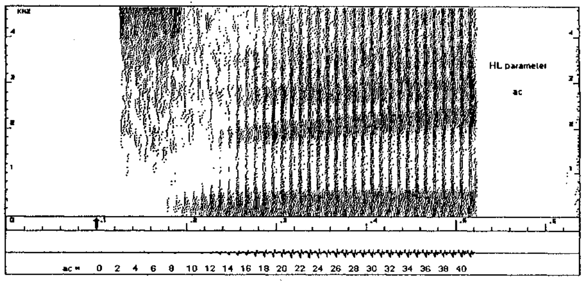

Fig. 3 - Spectrogram of sound produced by the synthesizer when the HL parameter ac *(area of vocal-tract constriction) is increased in 2 mm^2 steps from 0 to 40 mm^2. Formant frequencies correspond roughly to changes in vocal-tract shape.*

30 mm^2), the value of *f1* is determined by the size of the oral constriction, rising from 200 Hz at onset to 310 Hz at offset. Given some latitude for the trajectories of *f2* and *f3*, the example may be likened to an obstruent stop being released (in slow motion) into a following vowel. [The example has been arranged so that substantial pressure has built up in the system prior to the displayed waveform portion.] The figure shows that the increase in the parameter *ac* yields a substantial amount of unvoiced frication noise at onset (2 to 6), followed by voiced aspiration (8 to 12), various degrees of breathy voicing (14 to 24), and finally, modal voicing as the oral cavity becomes less and less constricted.

3.4 Voiced and voiceless stops

We examine next speech synthesis situations in which there are coincident changes in the values of the parameters *ag* and *ac*. Specifically, we consider the acoustic consequences of vocal-tract movements toward or away from complete closure at some point along its length while the glottis is exhibiting either a relatively adducted configuration or significant abduction.

Figure 4 shows some of the HL parameter patterns we have used to synthesize voiced and voiceless stops produced at an alveolar place of articulation (i.e., [d] and [t]). When a talker produces either stop, the tongue blade moves rapidly toward and away from the alveolar ridge during the closing and opening gestures, respectively. These movements are modelled in the synthesizer as changes in the parameter *ac*, which, in our simulations, achieves a maximum rate of change of about 5000 mm/s near the implosion and the release. These rapid movements of the tongue blade are reflected acoustically as abrupt formant transitions; they are simulated by particular patterns of change in the natural frequency parameter values (upper graph). Although these patterns are the same for [t] and [d], the direction and extent of change in the natural frequency parameter values will generally differ depending on the manner and place of the particular consonant.

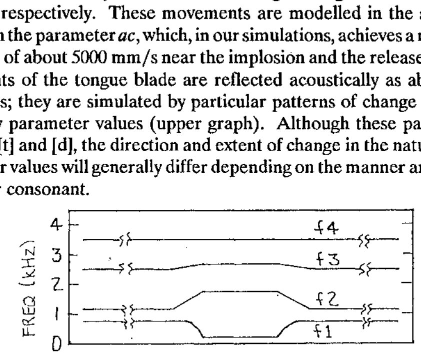

Fig. 4 - HL natural frequency (a) and orifice size (b) parameters used to synthesize the utterances [ada] and [ata].

For the voiced stop [d], modal voicing (*ag* = 4 mm^2) continues as the tongue blade moves to close off the vocal tract at the alveolar ridge, giving rise to a particular pattern of formant transitions. Of particular importance is the low-frequency transition of the first formant, which reflects the constricted state of the vocal tract and is a strong perceptual cue for voiced obstruents. When *ac* becomes zero, the calculated intraoral pressure increases, according to standard equations (e.g., [13][14]). If the aerodynamic conditions are appropriate (i.e., if the pressure across the glottis remains greater than about 3 cm H_2O) glottal vibration will continue throughout part of the closure. We introduce the HL parameter *ue* (not shown in the figure) to simulate the active expansion of the vocal tract that may be needed to maintain glottal vibration during closure.

Upon release of the stop at t_2, the pressure which has built up in the vocal tract during closure suddenly dissipates, giving rise to a short (5 to 10 ms) release burst. When the HL parameters *ac* and *ag* are controlled to follow the changes shown in Fig. 4, procedures within the mapping equations calculate the pressures and flows, including flow due to expansion of the vocal-tract walls and glottal surfaces in response to the increased intraoral pressure. From these pressures, flows, and areas, the time variations of the various LL acoustic source parameters are calculated, leading to estimates of AF (amplitude of frication noise), AH, and AV. To synthesize the voiceless aspirated stop, the parameter *ag* is increased during the closure interval for the stop, and is then decreased to a modal value about 50 to 100 ms following the release. As with the consonant [h], the mapping relations adjust various LL formant parameters as a consequence of the resistance and reactance of the glottal opening. For different rates of release and for different vowel environments, the aerodynamic conditions are such that the onset of vocal-fold vibration may be delayed by various amounts even though the time course of the control parameter *ag* is the same. Thus, for example, the voice onset time will automatically be greater for velars than for labials and will be greater for high-vowel contexts than for low-vowel contexts [5].

3.5 Voiced and voiceless fricatives

Figure 5 shows HL parameter settings *ac* and *ag* that we have used to synthesize the voiced and voiceless alveolar fricatives [z] and [s]. There are some similarities between the HL parameters for these utterances and those of the previous example. The transitions of the natural frequency parameters preceding and following times t_1 and t_2 (not shown in Fig. 5) are similar to those for the stop consonants, and again reflect movements of the tongue toward and away from the alveolar ridge. The patterns of variation for *ag* are somewhat different from those for the voiced and voiceless alveolar stops. Some increase in *ag* beyond its modal value during the voiced fricative is implemented in accordance with available data on airflow [12]. During the constricted interval for the voiceless fricative, there is a peak in *ag*, but this peak is centered near the middle of the fricative rather than near the release, as it is for an aspirated stop. These *ag* trajectories agree with those found in recent simulations of the aerodynamics of phonation in consonantal contexts [15].

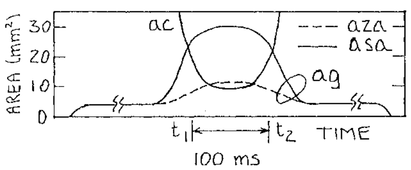

Fig. 5 - HL parameters ac *and* ag *used to synthesize [aza] and [asa]. The formant parameters are modified slightly from those shown in Fig. 4 for stop consonants.*

The primary difference between this example and the previous one is in the values that *ac* takes on over its time course. First, the maximum rates of decrease and increase of *ac* are slower for fricatives than for stops. Second, whereas the vocal tract is completely closed off in the production of stops, it is narrowly constricted for fricatives. Consequently the HL parameter *ac* lies in the vicinity of 8 to 10 mm^2 during the constricted interval for the fricatives.

Relative to stops, these differences in the patterns of change in *ag* and *ac* values yield significant changes in the resulting output. First, frication noise is generated near the tongue-blade constriction through most of the constriction interval. In the synthesizer, the amplitude of the frication source is calculated from the intraoral pressure and *ac*. Second, the intraoral pressure increase for the voiced fricative is not as great as it is for the voiced stop, so that glottal vibration can continue through the fricative more readily.

3.6 HL parameter *an*

Figure 6 shows the acoustic result of varying the parameter *an* from 0 to 40 mm^2 in steps of 3 mm^2 for two situations: (a) when there is an open vocal tract and modal voicing (e.g., for a nasalized vowel), and (b) when the vocal tract is closed (e.g., for a nasal sonorant). The natural-frequency parameters in both examples are set for a steady-state exemplar of the vowel /ɛ/ ($f1$ = 625, $f2$ = 1835, $f3$ = 2500). Varying the HL parameter *an* introduces an extra pole-zero pair into the cascade branch of the synthesizer, along with changes in the frequency and bandwidth of the first formant. When the nasal sidebranch is closed, FNP and FNZ are set to 500 Hz, cancelling one another. As *an* increases, the pole and zero take on frequency values that are determined by the value of *an*, the natural frequencies, and the value of *ac*. In the nasalized vowel, FNP moves up to a frequency of 900 Hz (and its bandwidth doubles), while FNZ rises to a frequency of 1050 Hz. The first formant frequency changes to 560 Hz while its bandwidth doubles. As the frequencies of FNP and FNZ rise, the zero increases at a faster rate, and crosses FNP near the middle of the left portion of the figure.

In the nasal sonorant example, all the values of the HL parameters are the

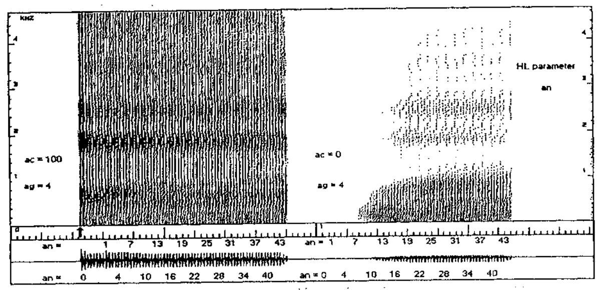

Fig. 6 - (a) Spectrogram of sound produced by the synthesizer when the HL parameter an *(area of velopharyngeal opening) during a vowel is increased in 3 mm^2 steps from 0 to 40 mm^2. (b) Same manipulation of* an *but with* ac *= 0 and formant frequencies adjusted to correspond to a closed vocal-tract configuration.*

same, except that the oral cavity is closed (*ac* = 0). The setting of *ac* = 0 causes the HL parameter *f1* to be about 200 Hz. When *an* = 0, or for small values of *an*, the intraoral pressure is such that vocal-fold vibration does not occur. For small *an*, some turbulence noise should be generated near the velopharyngeal opening. This noise is not incorporated in the model at present, since it is not believed to be perceptually salient for English. At larger values of *an*, voicing amplitude increases along with the frequency and bandwidth of the first formant, the frequencies of the nasal pole and zero rise to 930 Hz and 1560 Hz, respectively, and the tilt and the open quotient decrease progressively revealing the upper formants, as is evident in the spectrogram.

3.7 Nasal sonorant

Figure 7 shows values of the HL parameters *an* and *ac* for the utterance [ana]. Because the characteristics of vocal-tract closure and laryngeal adjustment are the same as for voiced stop in [ada], the settings of *ag* and *ac* are the same as those shown in Fig. 4. Timing of the velopharyngeal opening causes nasalization of the preceding and following vowel, and prevents the increase of intraoral pressure during the consonant closure interval.

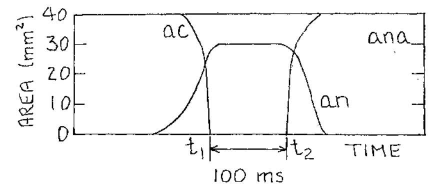

Fig. 7 - HL parameters ac *and* an *used to synthesize [ana]. Formant and* ag *parameters are the same as for [ada] in Fig. 4.*

Changes in the value of *an* are reflected by changes in the following LL parameters: the nasal pole and zero and their bandwidths (FNP, FNZ, BNP, and BNZ), and the frequency and bandwidth of the first formant (F1 and B1). Calculations of these LL parameters are effected in the mapping relations using established acoustic theory [1][8].

3.8 Place of articulation

In general, English consonants are generated using three different articulators: lips, tongue blade, and tongue body. In order to synthesize stop and nasal consonants with a particular place of articulation, we can expect that the HL parameter *ac* should have a prototypical time variation independent of whether the consonant is nasal, voiced, or voiceless. As we discussed earlier, the parameter *ac* for an alveolar closure will take a form like that shown in Fig. 4. The rate of closing and opening when *ac* is small is expected to be about the same independent of the vowel environment, and may be relatively independent of speaking rate. At times preceding the instant of closure or following the release by a few milliseconds, the area of the tongue blade constriction will be influenced by the preceding and following vowels. The duration t_2-t_1 is influenced by the speaking rate and by the relative stress of the adjacent syllables. Similar curves of *ac* versus time are expected for labial and velar consonants, except that the rates of opening and closing are different, especially for velars. Our initial experience with synthesis suggests that these rates for velars are about one-fourth of the rates for labials and alveolars. These rate differences are consistent with the limited articulatory, acoustic, and aerodynamic data that are available for stop consonants [9][10][16].

For all of the different places of articulation for stops and nasals, *f1* is low (at about 200 Hz) when the consonant is produced with a complete closure in the oral cavity [7]. Immediately preceding the implosion, *f1* falls from a value appropriate for the preceding vowel, and after the release *f1* increases. (The actual lowest natural frequency of the synthesizer, F1, may be somewhat higher than *f1* during the time there is a narrow constriction, as a consequence of an increased glottal opening or a velopharyngeal opening.)

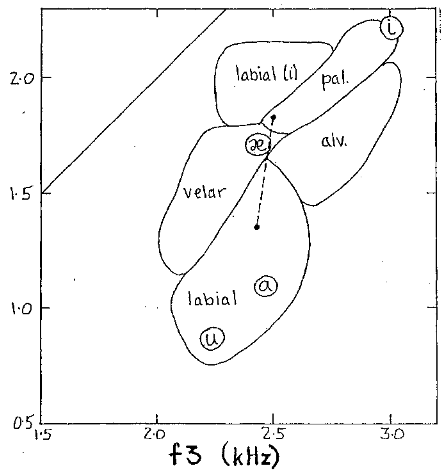

Fig. 8 - Natural frequencies f2 *and* f3 *of the vocal tract when a closure is made to produce various classes of stop consonants in English. Each labeled region delineates the range of values of* f2 *and* f3 *corresponding to the indicated stop closure when the consonant occurs in different vowel contexts. The upper labial region corresponds to a high -front vowel context. The circled symbols show* f2 *and* f3 *values for several vowels in English. The dashed line is the* f2-f3 *trajectory that is followed to synthesize the syllable /pɛ/. These regions are appropriate for an adult male vocal tract.*

The natural frequencies *f2* and *f3* during the closure interval for the stop consonant depend, of course, on where the closure is formed in the vocal tract. An *f2-f3* plot showing the regions that might be expected for velar, alveolar, and labial consonant closures of a male talker is given in Figure 8. These regions are tentative, and are based in part on our own measurements of formant transitions, in part on theoretical considerations, and in part on data available in the literature, such as those reported by Sussman [17]. In this plot, there are distinct ranges of *f2* and *f3* for labial, alveolar, and velar consonants. The labial and palatal-velar regions are divided into two parts depending on the following vowel. When a consonant closure is formed, the location of the point in the *f2-f3* plot corresponding to the closure depends to some extent upon the preceding and following vowels.

As an example, consider the labial consonant in the syllable /pɛ/. The *f2* and *f3* values corresponding to the labial closure are approximately 1350 and 2400 Hz, describing a point within the labial region of Fig. 8. The frequencies *f2* and *f3* then follow trajectories toward values for the vowel /ɛ/, as shown by the dashed line in the figure. As this movement in the *f2-f3* plane proceeds, the cross-sectional area *ac* of the labial opening increases, as does the lowest natural frequency, *f1*. When *ac* is small (less than about 30 mm^2), *f1* is strongly dependent on *ac*, and can be calculated approximately from acoustic theory. For small *ac*, then, *f1* cannot be specified independently, but is derived directly from *ac*. When *ac* becomes larger, *f1* is dependent on the entire vocal-tract shape, and can be specified more or less independently of *ac*.

We have observed above that a sequence of sources is generated in the synthesizer following the release of the consonant. Calculation of pressures and flows show that a frication source is produced at the labial constriction during the first few ms following release. This is followed by an aspiration noise source and then onset of glottal vibration. The aspiration source and the periodic glottal source are applied directly to the cascade branch of the formant synthesizer. The frication source, however, is filtered by adjusting the gains of components in a parallel array of formant resonators, together with a bypass path. The mapping equations contain tables of values for these gains depending on the values of *f2* and *f3* immediately following the consonant release. That is, these gains are determined by the position of points near the beginning of a trajectory like the one in Fig. 8. In the present example of a labial stop consonant, the only nonzero gain is the bypass path (parameter AB), because in the production of a labial stop there is no resonator anterior to the constriction to filter the source.

IV. SUMMARY

As the examples show, the HL extension to conventional formant synthesis both simplifies and broadens the task of synthesizing speech. Because the mapping relations incorporate the many constraints that exist between the LL parameters, the user of the synthesizer is freed from having to specify many acoustic details. There is, then, greater opportunity to concentrate on aspects of synthesis relating to the coordination between articulators and to the kinematics and dynamics of articulatory movements. [Work supported by the National Institutes of Health.]

V. REFERENCES

[1] Fant, G. *Acoustic Theory of Speech Production*. The Hague: Mouton 1960.

[2] Holmes, J. N. Influence of glottal waveform on the naturalness of speech from a parallel formant synthesizer. *IEEE Transactions on Audio and Electroacoustics, AU-21*, pp. 298-305, 1973.

[3] Klatt, D. H., and L. C. Klatt. Analysis, synthesis and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am. 87*, pp. 820-857, 1990.

[4] Fant, G. Speech communication research. *IVA (Royal Swedish Academy of Engineering Sciences) 24*, pp. 331-337, 1953.

[5] Williams, D. R., and C. A. Bickley. Examples of speech synthesis with high-level parameters controlling a Klatt synthesizer. *J. Acoust. Soc. Am. 91*, p. 2442, 1992. (Abstract)

[6] Stevens, K. N., and C. A. Bickley. Constraints among parameters simplify control of Klatt formant synthesizer. *J. Phonetics 19*, pp. 161-174, 1991.

[7] Fant, G. Vocal tract wall effects, losses, and resonance bandwidths. *Speech Transmission Laboratory Quarterly Progress and Status Report, 2-3*, Royal Institute of Technology, Stockholm, pp. 28-52, 1972.

[8] Stevens, K. N., G. Fant, and S. Hawkins. Some acoustical and perceptual correlates of nasal vowels. In R. Channon and L. Shockey (Eds.), *In honor of Lehiste: Ilse Lehiste Puhendusteos*. Dordrecht, Holland: Foris Publications. Pp. 241-254, 1987.

[9] Fujimura, O. Bilabial stops and nasal consonants: a motion picture study and its acoustical implications. *J. Speech and Hearing Res. 4*, pp. 233-247, 1961.

[10] Kent, R. D. and K. L. Moll. Cinefluorographic analyses of selected lingual consonants. *J. Speech and Hearing Res. 15*, 453-473, 1972.

[11] Perkell, J. S. *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. Cambridge, MA: MIT Press. 1969.

[12] Klatt, D. H., K. N. Stevens, and J. Mead. Studies of articulatory activity and airflow during speech. *Annals of the New York Academy of Sciences 155*, Article 1, pp. 42-55, 1966.

[13] Stevens, K. N. Airflow and turbulence noise in fricative and stop consonants: Static considerations. *J. Acoust. Soc. Am. 50*, 1180-1192, 1971.

[14] Rothenberg, M. The breath-stream dynamics of simple released-plosive production. *Bibliotheca Phonetica No. 6*. Basel: S. Karger. 1968.

[15] McGowan, R. S., L. L. Koenig, and A. Lofqvist. Aerodynamics of phonation: Effect of consonantal context on vowels. *J. Acoust. Soc. Am. 91*, p. 2421, 1992.

[16] Stevens, K. N. (in preparation) *Acoustic phonetics*.

[17] Sussman, H. M. The representation of stop consonants in three-dimensional acoustic space. *Phonetica 48*, pp. 18-31, 1991.