A Simple Program for Synthesizing Speech by Rules
Mark P. Haggard
Haskins Laboratories, New York City

Synthesis by rule is not new, having been performed about a decade ago in this laboratory. Any synthesis-by-rule scheme must owe something to the earlier work. Two factors, however, distinguish current attempts at synthesis by rule from that work. First, it is possible in general to produce high quality speech with a formant synthesizer and in particular to isolate prosodic from segmental information and to evaluate them independently. Second, the output of the rules, when written into a computer program for a digitally controlled synthesizer, is completely replicable and avoids the inaccuracies and the unconsciously intruding personal recipes of the operator.

It is obvious that the quality of the output of a set of rules should be evaluated along with some global measure of its complexity. Were this not so, the prize for the best speech output would always go to complex but trivial sets of rules, which were, in fact, closely related to natural utterances. Thus any scheme employing diphones, or specifying in detail transitions for every possible pair of phonemes cannot properly be called synthesis by rule. The simplicity criterion is also what underlies the above point about the intervention of the human operator. The individual's ability to synthesize from his experience better speech than his program puts out at a given stage means either that he has not fully developed or not unduly complicated his program. The achievement of high quality speech from rules, as by Holmes, Mattingly and Shearne (1964), suggests that it is not
too early to start on reducing the complexity of the rules.

The simplicity of the present set of rules rests upon the fundamental dichotomy in speech between vowels and consonants. The steady state portion of a consonant is given a unique categorical, spectral representation, but context-sensitive duration. With the exception of a few major allophonic variations, this unique spectrum is output for all occurrences of the consonant. The viability of this assumption rests on the poor discrimination of infra-phonemic differences in consonant sounds. Vowels, semi-vowels and diphthongs, on the other hand, are treated as highly context-sensitive, and are processed by various blending rules to give non-invariant transitions between consonant loci and vowel targets. Two possible sources of unnaturalness in synthesized vowels are avoided: there is no vowel steady-state between consonants of different loci and no actual reaching of vowel target between consonants with similar locus.

Since it is in the portions of speech synthesized while the tract is in vocalic mode that some speech quality is likely to be lost, this is what the current rules concentrate upon. The present system approximates the transition functions reasonably well. Research now planned will bring the formant frequency-amplitude and overall amplitude relationships under tighter control. This should improve voice quality.

The pragmatic value of the simplified scheme is as follows. By always extracting the transition from the vowel or semi-vowel, and by making only the transitions and vowels context-sensitive, a result not too different from that of Holmes et al. is achieved -- without providing the computational framework and the phoneme specifications required by their more general set
of rules. In other words, of the very large number of sets of rules that the Holmes' et al. framework could implement, it is believed that the good sets would resemble the present set in the important respects.

The rules are presented in detail in the Appendix. They involve chiefly contextual duration modifiers and transition calculators. To compare the output of this set with that of Holmes et al., two test sentences: "A bird in the hand is worth two in the bush" and "Could you have guessed that this speech came from rules?" were synthesized. In both sentences only intonation was taken from natural utterances. At this stage the computer program to implement the rules was not ready and the rules were followed slavishly by an operator typing the data in by hand.

The first sentence was evaluated and found to be reasonably intelligible, but not to have as good quality as the corresponding synthesis of Holmes et. al. One possible contribution to this difference is the durations, which were here calculated by simple rules, and in the corresponding Holmes et. al. version copied from real speech. To compare the current rules with the optimum of which the synthesis is capable, the rules output and a copy of a natural utterance were analyzed. The spectrograms in Fig.1 show (above) a transliteration of the "bird" utterance copied from the speech of John Holmes, by Ignatius Mattingly for the Haskins Parallel Formant Synthesizer and (below) the rules output\(^1\). Comparison indicates that the form of the transitions

\(^1\) The rules that produced this speech are denoted MK II and are a considerable improvement on MK I whose output was played by Mattingly to the 73rd Meeting of the Acoustical Society of America.
does not differ greatly and suggests that any poor quality should not be attributed to the assumptions underlying the transition forms.

Though the rules have not yet been perfected within the framework described above, they do have some desirable new features, e.g., the absence of a steady-state vowel in "bird." And considering the reasonable results so far and the total set's simplicity, further improvement by changing table values and adding detail seems to be justified. [See Appendix]

Appendix: Synthesis by rule of British English

(a) Synthesizer and Control

The formant synthesizer is described in outline elsewhere. [Haskins Laboratories Status Report on Speech Research (SR-1, 1965)] It is controlled by an Executive program written by Ignatius Mattingly, which accepts and punches paper tapes, accepts modifications to control parameters in store, and transmits control data to the synthesizer at a sampling rate under operator control. The control data digit assignments are given in Table 1.

(b) Tables of phonetic elements

These tables correspond to an acoustic specification of the ideal target values of the phonemes of RP English. Table 2 contains these specifications as the decimal values accepted by the Executive program for controlling the parallel formant synthesizer. As the formant frequency values are linearly spaced, the frequency corresponding to a given value N can be easily determined as follows: \( F_1 = (100 + 25N) \text{ cps} \); \( F_2 = (543 + 75N) \text{ cps} \); \( F_3 = (1190 + 167N) \text{ cps} \). The amplitude data require a separate key which is given in Table 3. A permanent staggering of the
Table 1
FORMANT SYNTHESIZER PARAMETER CONFIGURATION AND COMPUTER BIT ASSIGNMENT

Parallel Formant Synthesizer (DIFS)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Formant Frequencies</th>
<th>Formant Amplitudes</th>
<th>Excitation Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F₁</td>
<td>F₂</td>
<td>F₃</td>
</tr>
<tr>
<td>Bits per</td>
<td>5</td>
<td>5*</td>
<td>4</td>
</tr>
<tr>
<td>Sample</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>100 - 895 cps</td>
<td>545 - 2910 cps</td>
<td>1190 - 3700 cps</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 7 bits optional

Two computer words (24 bits each) are output for each time sample. Two of these are allocated to indexing, leaving 6 available for additional parameters such as a nasal selector.
<table>
<thead>
<tr>
<th>Element</th>
<th>$F_1$</th>
<th>$F_2$</th>
<th>$F_3$</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>$M$</th>
<th>$Ph$</th>
<th>$Fs$</th>
<th>$Dur$</th>
<th>$OA$</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>7</td>
<td>21</td>
<td>10</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>80</td>
<td>12</td>
</tr>
<tr>
<td>e</td>
<td>15</td>
<td>13</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>13</td>
</tr>
<tr>
<td>a</td>
<td>21</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>110</td>
<td>13</td>
</tr>
<tr>
<td>o</td>
<td>16</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>13</td>
</tr>
<tr>
<td>u</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>80</td>
<td>12</td>
</tr>
<tr>
<td>[0]</td>
<td>12</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>70</td>
<td>12</td>
</tr>
<tr>
<td>I</td>
<td>8</td>
<td>19</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>60</td>
<td>12</td>
</tr>
<tr>
<td>e</td>
<td>16</td>
<td>18</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>70</td>
<td>12</td>
</tr>
<tr>
<td>æ</td>
<td>20</td>
<td>15</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>70</td>
<td>12</td>
</tr>
<tr>
<td>ə</td>
<td>19</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>70</td>
<td>12</td>
</tr>
<tr>
<td>o</td>
<td>18</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>70</td>
<td>12</td>
</tr>
<tr>
<td>u</td>
<td>10</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>70</td>
<td>12</td>
</tr>
<tr>
<td>æ</td>
<td>16</td>
<td>13</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>I'</td>
<td>9</td>
<td>14</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>w</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>y</td>
<td>6</td>
<td>24</td>
<td>10</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>40</td>
<td>10</td>
</tr>
<tr>
<td>r</td>
<td>12</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td>l</td>
<td>7</td>
<td>10</td>
<td>12</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td>l'</td>
<td>11</td>
<td>5</td>
<td>11</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>100</td>
<td>10</td>
</tr>
</tbody>
</table>

| Bilabial locus | 5 | 5 | 7 | 4 | 2 | 0 | 0 | 2 | 0 | [60] | 8 |
| Alveolar locus | 7 | 16 | 10 | 0 | 2 | 5 | 3 | 2 | 3 | [60] | 8 |
| Velar locus | 8 | 20 | 7 | 0 | 4 | 5 | 1 | 2 | 1 | [70] | 8 |
| Velar locus' | 8 | 7 | 7 | 0 | 4 | 2 | 1 | 2 | 1 | [70] | 8 |
| Palatal locus | 7 | 20 | 7 | 0 | 6 | 7 | 3 | 2 | 1 | [60] | 6 |
| Labiodental locus | 7 | 10 | 9 | 0 | 1 | 1 | 0 | 1 | 0 | 60 | 10 |

5.7
Table 3

Amplitude conversion code for DIFS amplitude parameters. While 'intensity' is the usual term for parameters expressed as log power ratios, the term 'amplitude' is used by convention.

<table>
<thead>
<tr>
<th>All parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decimal</td>
</tr>
<tr>
<td>Figure</td>
</tr>
<tr>
<td>7  6  5  4  3  2  1  0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Decibel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equivalent</td>
</tr>
<tr>
<td>0  -3  -6  -9  -12  -15  -21  -∞</td>
</tr>
</tbody>
</table>

Base level of amplitude parameters re First Formant

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Formant Two</th>
<th>Formant Three</th>
<th>Formant Four</th>
<th>Fricative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-7db</td>
<td>-4.5db</td>
<td>-14db</td>
<td>-14db</td>
</tr>
</tbody>
</table>
amplitudes of the formants and fricative is necessary to use the number of bits available in the most efficient manner. Amplitudes relative to the first formant can be determined by simply adding -db values.

The phoneme table contains specifications for each element in terms of the eleven synthesis parameters, minus pitch. The remaining column (9) is allocated to element duration measured in milliseconds. With the exception of the values for nasal consonants /m,n,η/ these values were arrived at independently from those of Holmes et al. Parameter values for vowels and voiceless fricatives were found by judging isolated sounds produced by the synthesizer under manual control. Stop consonant and voiceless fricative data were taken from the original Haskins experiments on recognition cues (Ingemann, 1958) as were /w,r,y/. The /l/ allophones were determined in dynamic patterns synthesized under computer control and average values, seeming to fit most contexts, were derived. The spectral characteristics of voiceless stops and of affricates were determined in a similar manner, existing data on these sounds not being easily translatable to the present synthesizer. The duration data were determined partly by taking spectrograms, partly by inspection of duration data of Elaert (1964) on Swedish and partly by considerations of simplicity and symmetry, with the knowledge that these durations would be multiplied by various correction factors in syllables marked for stress, etc. The table is assumed to be substantially correct, although minor modifications are continually taking place, particularly in the data concerned with relative amplitudes. There is good overall agreement with the corresponding table of Holmes et al. but a few disagreements. To understand how the table
is operated upon by the rules, it is necessary to know the
details of some non-correspondences between the table and a
true phoneme inventory.

1. **Omissions from the table.**
   
   (a) **Diphthongs.** These are considered as single phonemes
       but unquestionably have a sound structure that requires postulating
       at least two participating phonetic elements. The present program
       therefore has no special table entries for diphthongs, only a
       special transition procedure for dealing with a pair of
       bracketed vowels. The procedure is the same for all diphthongs.

   (b) **The /h/ phoneme.** This phoneme is omitted from the
       table because it has no distinctive spectral structure. It
       adopts the formant frequency values of the following vowel while
       the excitation is that of a high pass filtered noise source at
       the glottis. It is handled by the rules as an override rule on
       the voicing control of the following vowel for 20 msec, or 40
       msec in the stressed case, extracted from the vowel duration.
       The same element is used when it is desired to express aspiration
       following a voiceless stop burst, which is normal in English for
       non-final position.

2. **Additions to the table.**

   (a) **The /o/ element.** RP English has no pure /o/ vowel,
       only the diphthong /ou/; however none of the other vowels in
       this region /ɔ, ɔ, u/ serves at all adequately as the initial
       element of that diphthong, hence the need for an /o/ element
       not corresponding to a phoneme.

   (b) **Multi-element phonemes.** Several types of consonant
       involve more than one row of table entries. For nasals and
       voiceless fricatives the steady-state is the second row of entries,
       and for affricates the third. Stop consonants and affricates also

5.10
have a steady-state portion consisting of silence (voiceless) or voice-bar (voiced) during consonant closure. As the spectral structure of this element hardly varies over all the sounds within these sub-classes, it comes not from the tables but from the rules. The specification of silence is obvious. A voice bar is a low-frequency, low amplitude first formant:

7 15 9 4 0 0 0 1 0 -- 4

The figure in the duration column of a burst element for these phonemes does not apply to the burst but to the duration of the silence or voice bar, which does differ slightly between the consonants in a sub-class (Elaert, 1964) although it is not known if this is significant perceptually. The burst portion is the second entry for each consoant where it occurs and has a standard duration of 10 msec.

3. **Major allophones entered in the table.**

Certain phonemes cannot be adequately characterized by one set of "target" values. These are:

(a) /l/. The /l/ allophone in final position is very different from that in initial position. The symbol /l'/ is used for the final allophone.

(b) /k, g, η/. The velars /k, g, η/ have loci that tend to vary continuously with the second formant frequency of the adjacent vowel. While this could be handled by a modification to the rules such a modification would have to operate on continuous parameters and would be unwieldy. As the locus is above the second formant for front vowels, (high second formant) and below it for back (low second formant) the variation can be expressed by a front and a back allophone of each of these consonants. This back allophone is denoted by a '.

(c) /s', f', θ'/ The spectral structure of fricatives
/s,ʃ,θ/ is affected by the rounding or lip closure of adjacent vowel /u, ,ou/, semi-vowel /w/ or consonant /p,b/. Rounded allophones /s',ʃ',θ'/ are provided.

(d) I'vowel. In British English the /I/ vowel has a special reduced form that is distinct from schwa. This special form is denoted I' and distinguished from schwa by a lower first formant.

4. Pause.

The element # is not usually considered a sound, though it is in a logical sense a phoneme. In the present program it serves three purposes: in the input to the program it signals a pause, so that certain duration modifications can take place. It also has a role in the output as a simple pause -- it has duration. Thirdly it functions like a vowel in transition computation: the excitation is low level aspiration which allows just audible release transitions in stop consonants in absolute final position. After a final vowel, a semivowel /l/ is not heard and simply represents the return of the speech apparatus to resting position.

(c) The rules.

The present rules are restricted to the production of temporal pattern of segmental information in speech from an input that is phonemic but includes certain extra symbols. The rules do not permit automatic derivation of the pitch parameters from grammatical markings; this aspect is ignored and the pitch is copied from a natural utterance in order to allow independent evaluation of the segmental properties of the speech from the prosodic properties. The examples which follow were produced by following the rules in punching the control data by hand (to be distinguished from copying a natural utterance), but the procedure for synthesis by rule is described in terms of the computer

5.12
program under development. The rules in the second half of the present program are metrically based, while those in the first part are mostly logically based.

(A) Logically based rules -- the first part of the program. In the first part of the program these comprise the duration modifiers, and are of 6 types. They are based on a simplification of some data of Elaert, (1964) and analysis of spectrograms of the speech of a few individuals. We expect to be able to refine and add to these rules considerably.

(1) Stress modifier. The beginning of the stressed syllable is marked; the end of the stressed syllable is located by a simple syllable delimitation procedure which divides intervocalic consonants equally between the two syllables or allocates the odd consonant to the following syllable in the case of an odd number. In the stressed syllable, all elements are lengthened except bursts; their durations are multiplied by a factor of 1.2. Stress also controls the portion of the vowel given aspiration as /h/ or post-stop aspiration.

(2) Prepausal modifier. Preceding the # sign the last vowel or both diphthong elements and all consonants following are lengthened. All vowels by a factor of 1.5 and consonants by 1.2. There is some indication that a sub-division of consonants should be made in this respect, but data on this have not yet been gathered.

(3) Pre-consonant modifier. All vowels and diphthong elements followed in the same syllable by a voiced consonant other than /l/ are lengthened by a multiplicative factor of 1.3. As voiced consonants are shorter than their voiceless cognates, this has some of the effect of holding syllable
duration constant.

(4) Cluster modifier. (a) General. All consonants and semivowels in clusters not divided by a stress mark or pause mark are multiplied in duration by 0.8. This rule can undoubtedly be refined. (b) Stop consonant cluster modifier. When stop consonants combine, the closure duration of the second is further decreased by 20 msec and the burst for the first is omitted. (Only in artificially careful real speech would it be seen).

(c) Homorganic stop rule. Closure duration of /t/ and /d/ following /s,n/, of /b/ and /p/ following /m/, and of /k/ and /g/ following /ŋ/ are further abbreviated by 20 msec.

(B) Metrically based rules -- the second part of the program. These operate upon data already in a quantitative form. They are chiefly designed to achieve a transition from one phone to another.

(1) Consonant transition rule. a) Consonant-vowel transition type 1: all consonants except semivowels /w,r,l,y/, have a table entry for the consonant locus. Transitions from vowel to locus and vice versa are worked out as follows.

For the first formant frequency a standard duration of 30 msec (3 time samples at normal sampling rate) is allocated to the transition. This proceeds from the vowel nucleus of the syllable out towards the consonant locus, and is taken entirely from the duration of the vowel (or semivowel) in the syllabic nucleus. For an initial transition the values for these three samples are 1/4, 1/2, and 3/4 of the way to the locus, vice versa for a final transition. If the vowel is shorter than 60 msec (/e/, /I'/) the outer transition portions survive at the expense of the center of the vowel, if the duration of the vowel is an odd number of samples. This linear interpolation procedure involves rounding off more often than it will not, but the 5.14
quantization interval is in the region of the just noticeable
difference for all parameters, so this is not a matter of
importance.

In the second and third formant frequencies and amplitudes,
a similar process takes place except that three sample values are
$1/3$, $1/2$ and $2/3$ of the way to the consonant locus. As in real
speech, the chief cue to the place of articulation of the
consonant, the second formant locus is a target which can be
inferred from the speech signal without being actually present in
it. Figure 2 illustrates this principle in two cases.
b) Consonant-vowel transition type two: Figure 2 also illustrates
the other type of transition observed chiefly in the second formant.
In real speech, final consonants are typically anticipated to an
extent where the final transition does not start from the vowel
target value. This reflects an inertial constraint upon the
forward-backward movement of the tongue mass, and is approximated
in the present rules by simply interpolating the second formant
between the $1/3$ way to locus points in the case of a consonant-
vowel-consonant syllable. This is not done when an initial
semivowel intervenes: in this case, or if there is only one
consonant in the syllable, interpolation type 2 reaches or
starts from the vowel or semivowel target value in the middle of
its connected element duration; with an even number of samples
the middle sample is defined as the one following the midpoint.
c) Consonant-to-consonant transitions: in real speech spectra
there are not many cases where such transitions are important.
One exception is entailed by the juxtaposition of fricatives and
lip-rounded or closed elements e.g. /-sp/-, /sw-/ etc. Although
there are transitions in the fricative noise parameters, the
Fig. 2 Schematic spectrogram as used, for example, on Pattern Playback synthesizer. The Figure shows how two aspects of real speech transition forms are approximated by linear interpolation. X represents the vowel target frequency of the second formant.
present program replaces them by a single rounded allophone, which may be all that perception requires. There is thus no need for transition rules to operate on the steady state portions of consonants.

(2) Vowel and semivowel transitions. Typically slower transitions are needed for these sounds than for the consonants described above. These sounds are classified together for transition purposes; the consonant transitions are all located within vowel or semivowel boundaries. Semivowels always combine with vowels and occasionally vowels combine transsyllabically. The program deals with them in the following way: the second and third formant frequencies and the overall amplitude are linearly interpolated between the mid-points of the adjacent elements (defined above). The first formant frequency undergoes a four-sample (40 msec) linear transition. The behavior of the second formant will allow the target values to be achieved in combinations of incompatible articulations like /iwi/ and uyu/, but this misrepresentation may not be crucial. It is probably of greater importance that this procedure will give a transition rate that is completely dependent upon speech rate. Experimental studies are planned to determine both the seriousness of this defect and the adequacy of the fixed-duration consonant transition model under 1 (a) for transitions over all frequency ranges.

(3) Diphthongs.
These are derived from elements already in table by a modification of the transition rules applied to both the second and first formant frequencies. Linear interpolation in these parameters proceeds from the third time sample at a point 1/3 of the way to initial consonant locus from the first diphthong element and reaches the target value of that element at the end of its corrected duration. If there is no initial consonant, the whole first element is a steady state. This gives the effect of
domination of the diphthong by the first element, which is appropriate. Interpolation is then continued from the beginning of the second element, (also corrected for duration by the same rules as vowels) to reach the target value of that element at the end of its duration. In the case of a final consonant it will reach the target at the fourth time sample from the end. These rules give reasonably good diphthongs. Transcriptions appropriate to their operation in RP English are /aɪ, ɔɪ, ɛɪ, ou, au, i, u, ʊ, ε/.

Transitions in amplitude parameters.

At the present there are no transitions in formant or fricative amplitude parameters. Thus the formant amplitude relationships just before reaching the consonant locus in /u/ would be those of /u/ rather than of a vowel in the /i/ region which would be more appropriate. It will be determined whether or not this is important and how precisely the constraints of real speech must be followed. Modifications are certainly possible. The amplitude transitions are at present handled by the overall amplitude parameter which takes on a value of 6, 8, and 10 or the same in reverse in the three time-samples of a type 1 transition. For a diphthong, a vowel-or semi-vowel-to-vowel transition, linear interpolation of overall amplitude is used between the element mid-points.

References

