AN ARTICULATORY SYNthesizer FOR PERCEPTUAL RESEARCH*

Philip Rubin, Thomas Baer, and Paul Mermelstein+

Abstract. A software articulatory synthesizer, based upon a model developed by Mermelstein (1973), has been implemented at Haskins Laboratories. The synthesizer is being used as a tool for studying the linguistically and perceptually significant aspects of articulatory events. A prominent feature of this system is that it easily permits modification of a limited set of key parameters that control the positions of the major articulators: the lips, jaw, tongue body, tongue tip, velum and hyoid bone. Time-varying control over vocal-tract shape and nasal coupling is possible by a straightforward procedure that is similar to key-frame animation: critical vocal-tract configurations are specified along with excitation and timing information. Articulation then proceeds on a directed path between these key frames within the time-script specified by the user. Such a procedure permits the required degree of control over articulator positions and movements. The organization of this system and its present and future applications are discussed.

INTRODUCTION

This paper provides a brief description of a software articulatory synthesizer implemented at Haskins Laboratories that is being used for a variety of experiments designed to explore the relationships between perception and production. A related paper, by Abramson, Nye, Henderson and Marshall (1979), describes in greater detail some of these experiments, focusing on the relationship between velar control and the distinction between oral stop consonants and their nasal counterparts. The intent of the present paper is to provide an overview of the actual design and operation of the synthesizer, with specific regard to its use as a tool for the perceptual evaluation of articulatory gestures.

The articulatory synthesizer embodies several sub-models. At its heart are simple models of six key articulators. The positions of these articula-

* Portions of this paper were presented at the 95th Meeting of the Acoustical Society of America, 16-19, May 1978, Providence, R. I.
+ Bell-Northern Research and INRS-Telecommunications, University of Quebec, Montreal, Canada.

Acknowledgment: The research reported here was supported under NSF Grant BNS-76-82023 and BRSQ Grant RR-05596 to Haskins Laboratories. We thank Steven E. Davis, Charles W. Marshall and Leonard Szubowicz for their significant contributions to the development of the articulatory synthesis program.

[HASKINS LABORATORIES: Status Report on Speech Research SR-57 (1979)]
tors determine the outline of the vocal tract in the midsagittal plane. From this outline the width function and, subsequently, the area function of the vocal tract are determined. Source information is specified at the acoustic, rather than at the articulatory level, and is independent of the articulatory model. Speech output during each frame is obtained after calculating, for a particular vocal-tract shape, the acoustic transfer function for both the glottal and fricative sources. For voiced sounds, the transfer function accounts for both the oral and nasal branches of the vocal tract. Continuous speech is obtained by a technique similar to key-frame animation (see below).

Although the synthesizer is capable of producing short segments of quite natural speech with a parsimonious input specification, its primary application is not based on these characteristics. The most important aspect of the synthesizer's design is that the articulatory model, though simple, captures the essential ingredients of real articulation. Thus, synthetic speech may be produced in which articulator positions or relative timing of articulatory gestures are precisely and systematically controlled, and the resulting acoustic output may be subjected to both analytical and perceptual analyses. Real talkers cannot, in general, produce utterances with systematic variations of an isolated articulatory variable. Further, for at least some articulatory variables—for example, velar elevation and the corresponding degree of velar port opening—simple variations in the articulatory parameter produce complex acoustic effects. Thus, the synthesizer can be used to perform investigations impossible with real talkers, or investigations that are difficult, at best, using acoustic synthesis techniques.

**THE ARTICULATORY MODEL**

The model that we are using was originally developed by Wermelstein (1973) and is designed to permit simple control over a selected set of key articulatory parameters. The particular set of parameters employed provides for an adequate description of the vocal-tract shape—while also incorporating both individual control over articulators and physiologically constrained interaction between articulators.

Figure 1 shows a midsagittal section of the vocal tract, with the six key articulators labeled: tongue body, velum, tongue tip, jaw, lips and hyoid bone position. (Hyoid bone position controls larynx height and pharynx width.) These articulators can be grouped into two major categories: **primary articulators**, whose movement is independent of other articulators (the jaw, velum and hyoid bone); and **secondary articulators**, whose positions are functions of the positions of other articulators (the tongue body, tongue tip and lips). The articulators of this second group all move relative to the jaw. In addition, the tongue tip moves relative to the tongue body. In this manner, individual gestures can be separated into components arising from the movement of several articulators. For example, the lip-opening gesture in the production of a /ba/ is a function of the movement of two articulators: the opening of the lips themselves, and the dropping of the jaw for the vowel articulation. Movements of the jaw and velum have one degree of freedom, all other articulators move with two degrees of freedom. **Movement of the velum** has two effects: It alters the shape of the oral branch of the vocal tract and, in addition, it modulates the size of the coupling port to the fixed nasal tract. This articulatory model is based, in large part, on knowledge of
KEY VOCAL TRACT PARAMETERS

C -- TONGUE BODY CENTER
V -- VELUM
T -- TONGUE TIP
J -- JAW
L -- LIPS
H -- HYOID

Figure 1
the anatomy and physiology and systematic observations of X-ray data and accompanying acoustic recordings of natural utterances (Mermelstein, 1973).

Figure 1 also shows the graphical display provided by the synthesizer system that permits the user to simply modify the midsagittal vocal-tract shape. To accomplish this, the user selects one of the six key parameters, moves a set of cross-hair cursors to specify its new position, and the new vocal-tract outline is immediately calculated and displayed on the graphics terminal. Before discussing further aspects of user interaction with the synthesizer system, an overview of the program structure will be presented.

PROGRAM ORGANIZATION

Figure 2 is a block diagram of the conceptual organization of the software for the articulatory synthesizer, as implemented on a DEC PDP-11/45 minicomputer. Input to the synthesizer is in the form of a list of positions of the key articulators that is arrived at in one of two ways. In what is called the manual mode of operation, input is derived from the static vocal-tract configuration that results from manual manipulation of the articulators as described above—a form of synthesis-by-art. Alternatively, input can be read from a previously stored table of values—synthesis-by-script. This second procedure, called the table mode of operation, will be discussed in more detail in the next section. (A third input mode, not shown in Figure 2, can also be used. In this mode an array of cross-sectional area values is specified as input, and earlier stages of the synthesis process are bypassed.) The articulatory input specifications are further supplemented by information about the source characteristics. These include, for the voiced source, input amplitude, fundamental frequency and two additional parameters that specify the properties of individual glottal pulses in either the time or frequency domain. For the frequency domain description these parameters are the pole frequencies. For the time domain description the two parameters specify the open quotient and speed quotient, using the pulse shape found most natural in a study by Rosenberg (1971). (The open quotient is the ratio of the duration of the glottal pulse to the duration of the whole glottal cycle. The speed quotient is the ratio of durations of the rising and falling phases of the pulse, and is thus a measure of skew.) In the case of fricative noise excitation, the amplitude and place of insertion of a pseudorandom noise source in the vocal tract are specified.

After the positions of the key articulators have been provided as input, either in manual or table mode, the program fleshes out this framework as a midsagittal section of the vocal tract, as was seen in Figure 1, and displays the shape, if desired. Cross-sectional areas are then calculated by superposing a grid structure on the vocal-tract outline and computing the intersecting points of the outline and the grid lines (Mermelstein, 1973). The center line of the vocal tract is determined as the line joining the mid-points of the grid segments subtended by the vocal-tract outline. This line length represents the length of the vocal tract modeled as a sequence of acoustic tubes of uniform cross-sectional areas. Vocal-tract widths are measured along lines perpendicular to the center line. Sagittal cross-sections are converted to cross-sectional areas with the aid of previously published data on the shape of the tract along its length. Different formulas are used for the pharyngeal region (Heinz & Stevens, 1964), oral region (Ladefoged, Anthony, & Riley,
STEPS IN ARTICULATORY SYNTHESIS

INPUT POSITION OF ARTICULATORS

VOCAL TRACT OUTLINE (DISPLAY)

CROSS-SECTIONAL AREAS (DISPLAY)

TRANSFER FUNCTION (DISPLAY)

SPEECH SYNTHESIS (OUTPUT)

DISPLAY WAVEFORM

DISPLAY PSEUDO-SPECTROGRAM

Figure 2
1971) and labial region (Mermelstein, Maeda, & Fujimura, 1971). The area function is then smoothed and approximated by a sequence of uniform tubes of fixed section length (.875 cm). To arrive at the cross-sectional area of the last section (nearest the lips), the tract is continued in a parabolic horn continuous in area with the computed cross-sectional area at the lips. This continuation allows truncation of the tract at a length value that is an integral multiple of the section length, and its termination by the appropriate acoustic impedance. Movements of the articulators, especially hyoid height and lip protrusion, affect the overall length of the vocal tract, resulting in a variable number of vocal-tract sections. When the determination of area values is completed, the vocal-tract transfer function is computed (Mermelstein, 1971, 1972), and displayed. (The mathematical theory for the transfer-function calculation is reviewed in Appendix A.) Speech output is then generated, at a sampling rate of 20 kHz, by feeding the glottal waveform through the digital filter representation of the transfer function.

Movements of the vocal tract are simulated using a quasi-static approximation. The positions of the articulators are determined, or specified, at the onset of every pitch period and the corresponding acoustic transfer function is computed. The resulting speech signal is obtained by concatenating the responses to individual pitch pulses of varying durations. Acoustic energy is usually propagated between pitch pulses, but may optionally be set to zero at the onset of every pulse. Output is generally produced within twenty to sixty times real time, which permits the kind of interactive use necessary for hypothesis-and-test research. Further, in the manual mode of control the user can obtain feedback at a number of different stages, in the form of displays of the vocal-tract outline, the cross-sectional area array and the acoustic transfer function. These varying forms of feedback information are extremely useful for providing the user with complementary descriptions of the particular articulatory shape being examined. In addition, they provide him with the opportunity to return to the manual adjustment stage if changes in the articulatory configuration are required.

Once an utterance has been completely generated, the speech signal produced can be played out, stored on a disk, or can be examined in more detail in a waveform editor program that is linked to the synthesizer. If desired, the signal can be compared with previously produced utterances or edited in a variety of ways. Also available as final output is a stylized spectrographic display, which serves as a summary statement of information about formant frequencies and their bandwidths, amplitude and fundamental frequency. In addition, an animated version of the synthesizer can be used for observing the dynamics of the entire articulatory sequence.

SYNTHESIS-BY-SCRIPT

The articulatory synthesizer, in its manual mode, provides an excellent means for examining vowel quality as a result of the excitation of the static vocal-tract shape. The use of the synthesizer in this mode, however, does not allow for the dynamic simulation necessary to model actual continuous speech. Therefore, another procedure has been implemented to provide for time-varying control over the movements of the articulators. The approach used is similar to what is called key-frame animation: the framework for a desired dynamic articulation is represented by a series of configurations of the vocal tract.
STEPS IN ARTICULATORY SYNTHESIS

INPUT POSITION OF ARTICULATORS

VOCAL TRACT OUTLINE
(DISPLAY)

CROSS-SECTIONAL AREAS
(DISPLAY)

TRANSFER FUNCTION
(DISPLAY)

SPEECH SYNTHESIS
(OUTPUT)

DISPLAY WAVEFORM

DISPLAY PSEUDO-SPECTROGRAM

Figure 2
1971) and labial region (Mermelstein, Maeda, & Fujimura, 1971). The area function is then smoothed and approximated by a sequence of uniform tubes of fixed section length (.875 cm). To arrive at the cross-sectional area of the last section (nearest the lips), the tract is continued in a parabolic horn continuous in area with the computed cross-sectional area at the lips. This continuation allows truncation of the tract at a length value that is an integral multiple of the section length, and its termination by the appropriate acoustic impedance. Movements of the articulators, especially hyoid height and lip protrusion, affect the overall length of the vocal tract, resulting in a variable number of vocal-tract sections. When the determination of area values is completed, the vocal-tract transfer function is computed (Mermelstein, 1971, 1972), and displayed. (The mathematical theory for the transfer-function calculation is reviewed in Appendix A.) Speech output is then generated, at a sampling rate of 20 kHz, by feeding the glottal waveform through the digital filter representation of the transfer function.

Movements of the vocal tract are simulated using a quasi-static approximation. The positions of the articulators are determined, or specified, at the onset of every pitch period and the corresponding acoustic transfer function is computed. The resulting speech signal is obtained by concatenating the responses to individual pitch pulses of varying durations. Acoustic energy is usually propagated between pitch pulses, but may optionally be set to zero at the onset of every pulse. Output is generally produced within twenty to sixty times real time, which permits the kind of interactive use necessary for hypothesis-and-test research. Further, in the manual mode of control the user can obtain feedback at a number of different stages, in the form of displays of the vocal-tract outline, the cross-sectional area array and the acoustic transfer function. These varying forms of feedback information are extremely useful for providing the user with complementary descriptions of the particular articulatory shape being examined. In addition, they provide him with the opportunity to return to the manual adjustment stage if changes in the articulatory configuration are required.

Once an utterance has been completely generated, the speech signal produced can be played out, stored on a disk, or can be examined in more detail in a waveform editor program that is linked to the synthesizer. If desired, the signal can be compared with previously produced utterances or edited in a variety of ways. Also available as final output is a stylized spectrographic display, which serves as a summary statement of information about formant frequencies and their bandwidths, amplitude and fundamental frequency. In addition, an animated version of the synthesizer can be used for observing the dynamics of the entire articulatory sequence.

SYNTHESIS-BY-SCRIPT

The articulatory synthesizer, in its manual mode, provides an excellent means for examining vowel quality as a result of the excitation of the static vocal-tract shape. The use of the synthesizer in this mode, however, does not allow for the dynamic simulation necessary to model actual continuous speech. Therefore, another procedure has been implemented to provide for time-varying control over the movements of the articulators. The approach used is similar to what is called key-frame animation: the framework for a desired dynamic articulation is represented by a series of configurations of the vocal tract.
The actual path of articulation is obtained by interpolating between these key frames. In essence, the user provides the synthesizer with a script, in the form of a table of values, for the complete articulation. Each line of the script consists of a "snapshot" of the vocal tract at some temporal point. The exact form of the articulation-over-time is then determined by linearly interpolating the articularatory parameters and computing the corresponding sequence of vocal-tract shapes.

An example of this procedure can be seen in Figure 3 for the synthesis of the utterance /da/. There are two "key" vocal-tract shapes specified as input. The first shape is an articulation appropriate for the onset of the production of the /da/, with the tongue tip occluding the vocal tract at the alveolar ridge. The vocal-tract shape appropriate for an /a/ is the second key configuration. The script, then, for this production begins with the first key shape specified at the onset of the utterance, at time 50 (ms), which permits a period of pre-release voicing. Release occurs at 50 ms, and all movement is completed by 120 ms, at which time the second key configuration is achieved. In this 70 ms period of rapid movement, a number of different intermediate shapes are calculated by linear interpolation, of which two (at 75 ms and 100 ms) are indicated in the figure. The production of the syllable continues for another 250 ms as specified by the time of the final /a/ configuration in the input. The additional specifications of this shape are necessary to indicate the changes in the excitation parameters, such as a fall-off in fundamental frequency towards the final third of the utterance (beginning at 240 ms), and a rapid decrease in the amplitude in the final 25 ms. The bottom half of Figure 3 shows the output generated from this articulation script in the form of a stylized spectrogram representing the first three formants, and the time-synchronized plots of fundamental frequency and overall output amplitude.

This straightforward procedure affords the user a flexible means of approximating productions observed in actual speech. Articulator movements are controlled by directing their path from shape to shape, with critical configurations serving as guides along the way. This allows for a simple specification of input information by the user. For example, the articulation for a /da/ can be represented, without refinements for naturalness, in the form of a script consisting merely of the two key vocal-tract shapes. Changes and comparisons between related articulations are easily accomplished. To produce a /na/ one can use the /da/ articulation previously described with the single modification of opening the velum to permit the required amount of nasalization. A series along the continuum from /da/ to /na/ can be created, then, by using ordered steps of velar opening, from a completely closed velum to one open to the degree desired for an acceptable /na/. Further, changes in timing relationships are also accomplished simply, by varying the time required to move between key configurations.

CONCLUSION

The design and implementation of the articulatory synthesizer is intended, as previously noted, to provide researchers with a flexible interactive tool for examining relationships between speech perception and production. Input parameters to the synthesizer are the positions of a limited set of major articulators and excitation and timing information. An important aspect
ARTICULATORY SYNTHESIS OF /da/

Figure 3
of the model's design is that speech sounds can be generated using controlled variations in timing and/or position parameters, and the output sounds used in formal perceptual tests. Another important aspect of the model is that the synthesis procedure is fast enough to make on-line interactive research practical.

One present application of the synthesizer is an investigation of detailed relationships between velar control and the perceptual oral-nasal distinction. Here, an important attribute of the synthesizer is its ability to produce complex variations in the acoustic output from a simple and natural variation of a single articulatory parameter—as contrasted with the more complicated procedures necessary to generate oral-nasal series by acoustic synthesis methods. In another application (Raphael, Bell-Berti, Collier, & Baer, in press), the articulatory synthesizer has been used to test hypotheses about articulation made on the basis of physiological (EMG) evidence on the one hand, and acoustic evidence on the other.

Additional future applications include a series of experiments intended to study the perceptual effects of variations in the relative timing of articulatory movements. Such investigations address the nature of the underlying organization of the speech act in terms of its dynamic "units." A planned technical improvement will be the development of a flexible display system that can function like a stop-frame projector. As we gain further insight into the anatomy and physiology of speech production, we would like to incorporate such additional knowledge into the model. The articulatory synthesis system, as described in this paper, already serves as a powerful research tool for examining perception—production relationships. We expect that the synthesizer's usefulness will grow as the system evolves and as we refine the issues to be investigated with its aid.

APPENDIX A

AREA-FUNCTION TO ACOUSTIC-TRANSFER-FUNCTION CALCULATION

The acoustic model of speech production, given the vocal tract area function, is indicated in Figure 4. The various branches of the vocal tract are treated as linear two-port networks. For voiced or aspirated sounds (Fig. 4a), where the velar port may be partially open, the glottal source \( U_g \), feeds the left part of the pharyngeal branch. (For convenience, the glottal source impedance has been brought inside the box.) The right side of the pharyngeal branch is connected in parallel with the nasal and oral branches. On the right side of these two boxes appear the radiated nasal and oral pressure, each across an open circuit, since radiation characteristics have been brought inside the boxes. The output sound is the sum of the nasal and oral pressures. The junction point, at which the three subsystems are connected in parallel, corresponds anatomically to the top of the pharynx, at the level of the velopharyngeal port. However, if the nasal port is closed, the nasal branch drops out, and it no longer matters at what anatomical level the two remaining boxes split.

For fricative sounds (Fig. 4b), there is a noise source anterior to a constriction. The system splits into two parts: a front cavity, and a back cavity (which anatomically includes the constriction and also includes the
Figure 4a: Block diagram for voiced and aspirated sounds

Figure 4b: Block diagram for fricatives
source resistance associated with the noise). Across the other side of the front cavity is the radiated pressure from the mouth, where, again, the radiation characteristics have been brought inside the box. Across the other port of the back cavity appears the glottal source, if any. As before, glottal impedance has been brought inside the box.

In both parts of Figure 4, the glottal source and the leftmost box can be replaced by their Norton equivalent. If the two-port network obeys reciprocity, the Norton equivalent source, \( U_{\text{geff}} \), is related to the actual glottal source, \( U_g \), by the relation

\[
U_{\text{geff}} = U_g G_p ,
\]

where \( G_p \) is the open-circuit pressure gain \( p_g/p_p \mid U_g=0 \).

Using the Norton equivalent for the pharyngeal branch, as indicated in Fig. 5, it can be seen that the output for Fig. 4a is

\[
P_m + P_n = \frac{U_{\text{geff}}}{1/Z_m + 1/Z_n + 1/Z_p} (G_m + G_n) ,
\]

where \( G_m \) and \( G_n \) are the open-circuit gains \( p_m/p_p \mid U_g=0 \) and \( p_n/p_p \mid U_n=0 \), respectively. Therefore the transfer function is

\[
\frac{P_m + P_n}{U_g} = \frac{G_p (G_m + G_n)}{1/Z_m + 1/Z_n + 1/Z_p} .
\]

(1)

If the nasal tract is not present (that is, if the velopharyngeal port is closed), then \( Z_n=0 \) and \( G_n=0 \). Then, a corresponding equation accounts for the glottal component in Fig. 4b.

The transfer function for the fricative component in Fig. 4b is

\[
\frac{P_m}{P_S} = \frac{Z_m}{Z_m + Z_p} G_m .
\]

(2)

Thus, all the relevant transfer functions can be calculated if the input impedances looking from the junction point and the open-circuit pressure gain functions of all branches of the vocal tract are known. An iterative procedure for calculating these functions is described below.

For the purpose of calculations, the vocal tract is modeled as a series of uniform tubes with varying cross-dimensional but uniform length of 0.875 cm. A plane wave entering one end of such a section reaches the other end with a half time-unit (0.025 msec.) delay and an attenuation \( a^{1/2} \), which depends on the cross-sectional area. We will consider one such section with cross-sectional area \( A \), looking into an acoustic impedance \( Z_L \) from one end. When seen from inside the tube, this impedance produces a complex reflection coefficient.
Figure 5: Equivalent block diagram for voiced and aspirated sounds
\[
\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0},
\]
(3)

where \( Z_0 = 40/A \) is the characteristic impedance of the tube. (All physical quantities are expressed in cgs units.) The impedance, \( Z \), looking into the other end of the tube is then

\[
Z = Z_0 \frac{1 + z^{-1} \Gamma}{1 - z^{-1} \Gamma},
\]

where \( z \) is the Z-transform variable and \( \Gamma \) is expressed in Z-transform notation. The pressure gain across the tube is

\[
G = (\alpha^{1/2}z^{-1/2}) \frac{1 + \Gamma}{1 + \alpha z^{-1} \Gamma}.
\]

Consider now tube section \( n \), of area \( A_n \), looking into a load impedance \( Z_{n-1} \), which produces the reflection coefficient \( \Gamma_n \). The next section, which has area \( A_{n+1} \), sees an acoustic impedance

\[
Z_n = \left(\frac{40}{A_n}\right) \frac{1 + \alpha_n z^{-1} \Gamma_n}{1 - \alpha_n z^{-1} \Gamma_n},
\]

which can be considered a reflection coefficient

\[
\Gamma_{n+1} = \frac{r_n + \alpha_n z^{-1} \Gamma_n}{1 + r_n \alpha_n z^{-1} \Gamma_n},
\]

where

\[
r_n = \frac{A_{n+1} - A_n}{A_{n+1} + A_n}.
\]

This can, in turn, be used to find the impedance or reflection coefficient on the other side of section \( n+1 \), and the gain across it.

We now express the reflection coefficients as ratios of polynomials, so that

\[
\Gamma_n = \frac{P_n}{Q_n},
\]

where \( P \) and \( Q \) are polynomials in \( z \). Therefore,

\[
P_{n+1} = r_n Q_n + \alpha_n z^{-1} P_n
\]

and

\[
Q_{n+1} = Q_n + r_n \alpha_n z^{-1} P_n.
\]

(4)

(5)
The impedance into section \( n \) from the end of section \( n+1 \) is

\[
Z_n = \left(40/A_{n+1}\right) \frac{Q_{n+1} + P_{n+1}}{Q_{n+1} - P_{n+1}},
\]

and the pressure gain across section \( n \) is

\[
G_n = (a_n^{1/2}z^{-1/2}) \frac{Q_n + P_n}{Q_n + a_n^{-1}z^{-1}P_n}.
\]

But

\[
Q_n + a_n^{-1}z^{-1}P_n = \frac{Q_{n+1} + P_{n+1}}{1 + r_n},
\]

so that

\[
G_n = a_n^{1/2}z^{-1/2}(1 + r_n) \frac{Q_n + P_n}{Q_{n+1} + P_{n+1}},
\]

and the gain over sections 1 to \( N \) can be calculated:

\[
G = z^{-N/2} \frac{Q_1 + P_1}{Q_{N+1} + P_{N+1}} \prod_{n=1}^{N} (a_n^{1/2}(1 + r_n)).
\]

To begin the transfer-function calculation, the source impedance or the radiation impedance (and gain) at the end of each branch of the vocal tract must be known and expressed as a ratio of polynomials in \( z \). These are then used to determine the \( Q \) and \( P \) polynomials at the external end of the branch, using equation 3, and the iterative equations 4 and 5 are applied one section at a time until \( P \) and \( Q \) at the other (internal) end of the branch are determined. Lumped losses can also be introduced during these iterations. Both the impedance and gain can then be calculated, using equations 6 and 7, respectively. When this is done for all branches of the vocal tract, the glottal and fricative transfer functions can be calculated, using equations 1 and 2. Standard techniques are then used to implement the transfer functions as digital filters to perform the synthesis.

REFERENCES


Ladefoged, P., Anthony, J., & Riley, D. Direct measurements of the vocal


