

Nov. 24, 1970

J. L. FLANAGAN
AUTOMATIC GENERATION OF VOICELESS EXCITATION IN
A VOCAL-TRACT SYNTHESIZER

3,542,955

Filed April 29, 1968

2 Sheets-Sheet 1

FIG. 1

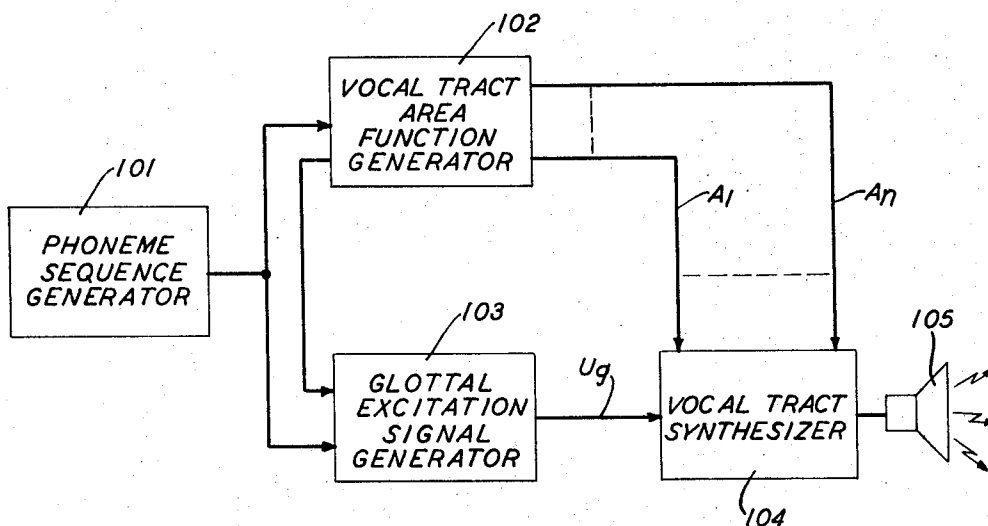
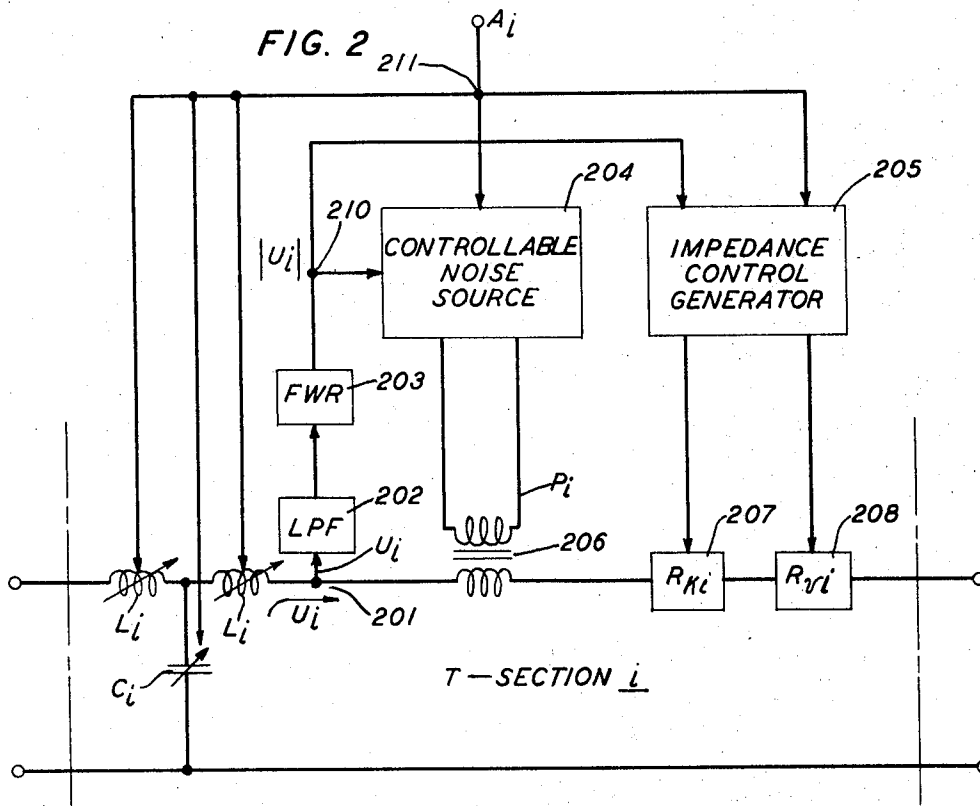


FIG. 2



INVENTOR
J. L. FLANAGAN
BY *G. E. Hurich Jr.*
ATTORNEY

Nov. 24, 1970

J. L. FLANAGAN
AUTOMATIC GENERATION OF VOICELESS EXCITATION IN
A VOCAL-TRACT SYNTHESIZER

3,542,955

Filed April 29, 1968

2 Sheets-Sheet 2

FIG. 3

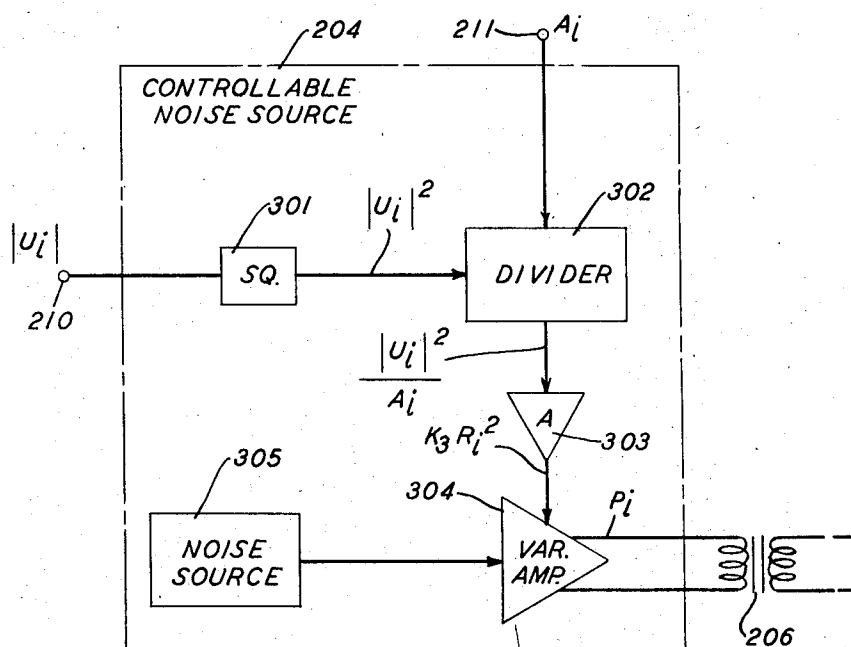
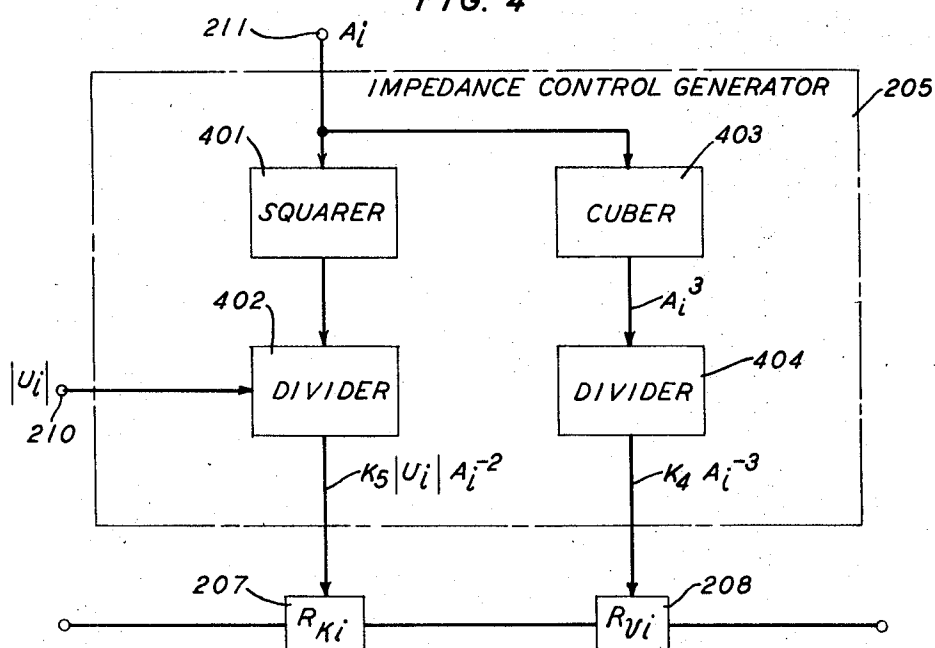


FIG. 4



1

3,542,955

AUTOMATIC GENERATION OF VOICE- LESS EXCITATION IN A VOCAL-TRACT SYNTHESIZER

James L. Flanagan, Warren Township, Somerset County,
N.J., assignor to Bell Telephone Laboratories, Incorporated,
Murray Hill, N.J., a corporation of New York

Filed Apr. 29, 1968, Ser. No. 724,931

Int. Cl. G101 1/08

U.S. Cl. 179—1

8 Claims 10

ABSTRACT OF THE DISCLOSURE

Natural-sounding synthesized speech is produced in an analog vocal-tract synthesizer by selectively inserting a noise signal, in addition to a glottal excitation signal, into each section of the synthesizer. The magnitude of the noise signal and the internal resistance of the noise source are automatically controlled in response to the current flowing in and the area control signal applied to each synthesizer section.

BACKGROUND OF THE INVENTION

This invention relates to speech synthesis and, more particularly, to apparatus for exciting an analog speech synthesizer to generate synthesized speech sounds.

Several systems have been proposed which attempt to produce natural-sounding speech through synthesis. In one of the most promising of these systems an attempt is made to duplicate the transmission properties of the human vocal tract and to supply suitable vocal excitation to a model of the vocal tract constructed to exhibit these properties. Since the exact process by which human speech is generated is not completely understood, speech produced in such a synthesis system is often unnatural sounding. Difficulties have been experienced in duplicating signals representative both of the voiced and unvoiced portions of human speech. Fortunately, it is now possible to develop voiced speech portions of a very high quality. The improvement, as described in my copending application, J. L. Flanagan, filed Aug. 29, 1967, Ser. No. 664,130, is achieved by further simulating the glottal section of the human articulatory system as a second order oscillatory system. The oscillatory system is controlled in accordance with physiological parameters of subglottal pressure, vocal cord tension, and vocal tract shape. Although this system greatly improves the quality of voiced sounds, it does not provide sufficient control of the synthesizer to achieve a like improvement in unvoiced sounds.

Generation of natural-sounding speech also requires that the synthesizer be properly excited in the voiceless domain. Voiceless excitation is of two types, namely, continuant (fricative) and transient (stop). Fricative excitation is essential for generation of sounds, such as /s/ and /sh/. Fricatives are generated in the human articulatory system by turbulent air flow (noise) at a constriction in the vocal tract. At some given velocity, the flow through the constriction becomes turbulent and thereby constitutes a localized random noise source. The location of the effective noise source and the magnitude of the noise, however, are difficult to determine.

2

Stop excitation is produced by making a complete closure of the vocal tract. Pressure is built up behind the closure and, when released, a step function of pressure is applied to the vocal tract. Generally, stop excitation is followed by some fricative excitation (or aspiration) caused by turbulence generated after the release of the pressure. The position of the stop source, i.e., the step function, and the magnitude of the signal are also difficult to determine.

A system which attempts to reproduce speech, in response to an analysis of actual human speech, is disclosed in U.S. Pat. 3,042,748 issued to G. Rosen, July 3, 1962. Rosen employs apparatus for inserting signals into an analog synthesizer to produce synthetic speech sounds in response to a voiced signal. Rosen, however, attempts to control externally the magnitude of and place of insertion of the signals. Signals so generated yield speech sounds having poor quality. In such a system, difficulties are experienced in locating where the noise signals are to be inserted for obtaining satisfactory synthesized sounds. Although such a system may be useful in certain applications, it is unsatisfactory for practical use in the synthesis of human speech in response to coded signals, i.e., to the synthesis of speech by rule.

Therefore, it is an object of this invention to generate natural-sounding speech sounds.

Another object of this invention is to generate voiceless excitation in a vocal-tract synthesizer in response to signals present within the synthesizer.

SUMMARY OF THE INVENTION

In accordance with this invention, these and other objects are accomplished by turning to account the flow characteristics of the human articulatory system. It is important to recognize that voiceless sounds are produced in the human articulatory system by the flow of turbulent air through a constriction in the vocal tract, or by a closure of the vocal tract. Thus, in an electrical counterpart to the human system, controllable signals developed to represent the momentary changes in air flow caused by variation in a vocal-tract configuration may be used to control voiceless excitation in a synthesizer to produce natural-sounding artificial speech.

It is known that flow through a constriction may be characterized by Reynold's number, and that Reynold's number is proportional to the particle velocity of the flow through the constriction and to the width of the constriction. Furthermore, it is known that noise pressure at a fixed distance in front of the mouth and, in turn, the noise pressure at a given location, for example, at a constriction within the vocal tract, is proportional to the square of Reynold's number in excess of a critical threshold level.

Thus, according to the invention, voiceless excitation is developed in a vocal-tract synthesizer by use of variable random energy (noise) sources, each having controllable internal impedances. The noise is selectively inserted in each of a plurality of sections, which make up the synthesizer. Control of the noise source and its impedance is effected by using signals present within each synthesizer sections, namely, area control signals and current level.

Accordingly, in the practice of this invention, voiceless excitation is developed in each section of a vocal-tract synthesizer by sampling the current flowing in that sec-

tion and utilizing a signal representative of the magnitude of the current in conjunction with the area control signals already supplied to the section to generate signals representative of the voiceless portion of speech.

Therefore, a feature of the present invention is the generation of signals in response to instantaneous changes in the shape of and the flow within a vocal-tract synthesizer section to develop voiceless excitation for generating synthetic speech.

Another feature of the invention permits voiceless excitation signals to be used in conjunction with glottal excitation signals to form voiced fricative synthetic speech sounds.

These and other objects and advantages of the invention will be more fully understood from the following detailed description of an illustrative embodiment thereof taken in connection with the appended drawings.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 shows in block schematic form a speech synthesis system in which the present invention may be utilized;

FIG. 2 illustrates a section of an analog synthesizer that employs the present invention;

FIG. 3 shows in block schematic form the details of the controllable noise source of FIG. 2; and

FIG. 4 shows in block schematic form the details of the impedance control generator of FIG. 2.

DETAILED DESCRIPTION OF THE INVENTION

A spoken word is composed of a sequence of linguistic elements called phonemes. Generally, to produce synthetic speech, phonemes, or sequences of phonemes, together with suitable stressing must be specified in a manner directly acceptable by a synthesizer. The synthesizer must be excited by signals representative of these phonemes, in both the voiced and unvoiced domain, in order to generate natural-sounding speech sounds.

FIG. 1 shows in simplified block form a synthesis by rule system suitable for generating natural-sounding speech in response to coded signals. Phoneme sequence generator 101 is utilized to produce appropriate signals for initiating the synthesis of speech sounds. Upon request, this information is supplied to vocal-tract area function generator 102 and to glottal excitation signal generator 103. Area control signals A_1 through A_n for controlling the synthesis of voiced sounds by transmission line vocal-tract synthesizer 104 are generated by vocal-tract area function generator 102. Synthesizer 104 also requires suitable excitation, U_g , which is produced in glottal excitation generator 103. Synthesizer 104 responds to the applied control signals and develops artificial speech sounds which may be used to energize a transducer such as loud speaker 105. A synthesis system such as this is described in detail in my copending applicator (Ser. No. 664,130, filed Aug. 29, 1967) cited above. Although that system generates signals which are adequate to excite vocal-tract synthesizer 104 for voiced sounds, it does not afford sufficient control over synthesizer 104 to generate high-quality unvoiced sounds.

In order to generate high-quality synthetic speech, a synthesizer must be capable of automatically responding to instantaneous changes in the shape of the vocal tract and to changes in flow within the vocal tract. Thus, in order to simulate the human vocal tract, speech synthesizer 104 may comprise n sections, each of which represents a portion of the human vocal tract. To excite vocal-tract synthesizer 104 for generating unvoiced and voiced fricative sounds, in addition to voiced sounds, noise signals must be inserted into each of the synthesizer sections. The inserted noise signals must be representative of noise generated by turbulent flow within the human vocal tract.

It is known that flow through a constriction may be characterized by Reynold's number (R). Thus, for a typi-

cal section, for example section i of n section synthesizer 104

$$R_i = \frac{u_i W_i \rho}{\mu} \quad (1)$$

where:

u_i is the acoustic particle velocity flowing in synthesizer section i ;

W_i is the effective width of synthesizer section i ;

ρ is the air density; and

μ is the coefficient of viscosity.

Current flowing in a section of a synthesizer is representative of acoustic volume velocity U_i . Thus, since U_i is equal to the product of the particle velocity u_i and the area A_i and since μ and ρ are generally constant, Equation 1 may be expressed as:

$$R_i = K_1 \frac{|U_i|}{\sqrt{A_i}} \quad (2)$$

Experimental measurements have indicated that noise pressure at a fixed distance in front of the mouth and, in turn, the pressure at a given location, e.g., at a constriction, within the vocal tract, is proportional to the square of Reynold's number in excess of a critical threshold level. That is,

$$P_i = K_2 (R_i^2 - R_c^2) \quad (3)$$

where $R_i \geq R_c$ and K_2 is a constant.

In a practical system, because of the squaring, the threshold term R_c is not critical and actually may be neglected. Thus, Equation 2 can be simplified to

$$P_i = K_2 R_i^2 \quad (4)$$

By simple substitution, Equation 4 reduces to

$$P_i = K_3 \frac{|U_i|^2}{A_i} \quad (5)$$

Therefore, a localized source of noise may be controlled by signals already present within the synthesizer. For example, the area control signal A_i and the current flowing within the section which is representative of U_i .

Furthermore, experimentation has indicated that the internal impedance of the localized noise source may be approximated by a series connection of variable resistive elements, namely, R_{xi} and R_{vi} .

$$R_{xi} = K_5 |U_i| A_i^{-2} \quad (6)$$

$$R_{vi} = K_4 A_i^{-3} \quad (7)$$

These terms are described in greater detail in section 3.52 of my book, "Speech Analysis, Synthesis and Perception," Academic Press, 1965, in connection with the vocal cord orifice model.

Thus, FIG. 2 shows in schematic form section i of vocal-tract synthesizer 104 comprising a plurality of n sections which respond to signals present within a vocal-tract model to excite the synthesizer for generating the unvoiced, i.e., fricative and stop, portion of synthesized speech sounds. Inductors L_1 and capacitor C_1 are primary impedance elements which form one T-section of transmission-line type vocal-tract synthesizer 104. The component values of inductors L_1 and capacitor C_1 are controlled in response to vocal-tract area signal A_i . A synthesizer T-section of this type and apparatus for controlling the T-section components is described in greater detail in the Rosen patent cited above.

Current U_i is developed in T-section i in response to the application of glottal excitation signal U_g (FIG. 1) to vocal-tract synthesizer 104. Thus, the resultant current U_i flowing within section i is sampled at 201. Current U_i normally comprises a D.C. component and higher spectral components. Generally, the D.C. component is several times greater in magnitude than the higher spectral components. Since the magnitude of the noise signal to be inserted in the synthesizer section and the value of

the internal impedance of the noise source are dependent upon the magnitude of U_i , the system may tend to be unstable. Thus, further to stabilize the system, U_i is filtered via low-pass filter 202 to eliminate the higher spectral components, which are generally responsible for such instability. Moreover, filter 202 minimizes the possibility of feeding back high spectral components that are inserted into the synthesizer by noise source 204, or by the noise sources of any of the other sections of the synthesizer, thereby further insuring stability of the system. Experience indicates that a low-pass filter having a cut-off frequency of approximately 500 Hz. is sufficient for this purpose. After filtering, current signal U_i is applied to full wave rectifier circuit 203 to obtain its absolute value $|U_i|$. Obviously, any other arrangement for obtaining the absolute value of U_i may be utilized as desired.

A signal representative of $|U_i|$ is thereafter applied to controllable noise source 204, to be discussed in greater detail in connection with FIG. 3, and to impedance control generator 205, to be discussed in connection with FIG. 4. Area control signal A_i is also applied to controllable noise source 204 and to impedance control generator 205. Signal A_i is utilized in conjunction with a signal representative of $|U_i|$ to generate the desired signal P_i , representative of noise, and to generate signals for controlling the resistive values of impedances R_{ki} and R_{vi} . Noise signal P_i is selectively inserted into synthesizer section i via transformer 206. Thus, the noise signal inserted and the resistive values of R_{ki} and R_{vi} cooperate to vary the flow of current U_i in the synthesizer section, thus to generate natural-sounding speech sounds.

FIG. 3 shows in block schematic form details of controllable noise source 204. Signals representative of A_i and $|U_i|$ are supplied, in accordance with the invention, to controllable noise source 204 for generating a noise signal P_i , in accordance with Equation 5. Accordingly, $|U_i|$ is squared in squaring network 301, and thereafter applied to one input of dividing network 302. A_i is applied to another input of dividing network 302. The output from divider 302 is applied to amplifier 303, where its amplitude is adjusted. The output from amplifier 303 is thereafter applied to control the gain of variable gain amplifier 304. A noise signal generated in noise source 305, which may be, for example, a gas tube, is applied to the input of variable gain amplifier 304. The output signal P_i of amplifier 304 is applied to transformer 206 for insertion into the synthesizer section.

A block schematic diagram of impedance control generator 205 is shown in FIG. 4. Signals A_i and $|U_i|$ are supplied to generator 205, wherein they are utilized to generate signals for controlling the impedance values of R_{ki} and R_{vi} , in accordance with Equation 6 and 7, respectively. Thus, to generate a signal for controlling the impedance value of R_{ki} , A_i is applied to squaring network 401. The output of squaring network of 401, which represents A_i^2 , is applied to one input of dividing network 402. A signal representative of $|U_i|$ is applied to a second input of dividing network 402. The output signal, $(K_5|U_i|A_i^{-2})$, from dividing network 402 is representative of Equation 6 and is applied to variable resistor 207, which may be either a rheostat driven by a servomechanism, an appropriately biased field effect transistor, or the like. Signal A_i is also applied for cubing to network 403, where it is converted to A_i^3 . Thereafter A_i^3 is divided and appropriately scaled in network 404. The output from network 404 is representative of Equation 7. This signal, $(K_4A_i^{-3})$, is applied to variable resistor 208.

The above-described arrangements are, of course, merely illustrative of the application of the principles of this invention. Numerous other arrangements may be devised by those skilled in the art without departing from the spirit or scope of the invention.

What is claimed is:

1. In a section of a vocal-tract synthesizer having a

plurality of sections each responsive to an individual vocal-tract area function signal, apparatus which comprises:

means responsive to said vocal-tract area function signal and to a signal representative of the magnitude of a volume velocity flow signal developed in a circuit path of said synthesizer section for generating a set of control signals;

controllable means responsive to at least one of said control signals for generating a random energy signal having a magnitude controlled in accordance with a preestablished relationship with said control signals; and

means for inserting said energy signal into said circuit path of said synthesizer section, thereby to develop voiceless excitation.

2. Apparatus as defined in claim 1, wherein said control signal generator includes:

means responsive to said flow signal for generating a signal representative of the magnitude of said flow signal squared; and

means responsive to said squared signal and said area function signal for generating a signal representative of the magnitude of random energy to be inserted into said circuit path of said synthesizer section.

3. Apparatus as defined in claim 1, further including controllable impedance elements connected in series with said inserting means, said impedance elements being responsive to others of said control signals for varying the magnitude of the impedance in said circuit path of said synthesizer section in accordance with a preestablished function of said control signals.

4. Apparatus as defined in claim 3, wherein said control signal generator includes:

means responsive to said local-tract area function signal for generating a signal to control the impedance value of one of said impedance elements; and

means responsive to the magnitude of said volume velocity flow signal and to said area function signal for generating a signal for controlling the impedance value of another of said impedance elements.

5. In an analog vocal-tract synthesizer having a plurality of sections, each responsive to an individual vocal-tract area function signal, apparatus for developing voiceless excitation in each of said sections, which comprises:

a controllable source of noise energy having variable internal impedance elements;

means for generating a signal representative of the magnitude of a current developed in a circuit path of said synthesizer section, said current signal being representative of the volume velocity flow in said synthesizer section;

means responsive to said vocal-tract area function signal and to said current magnitude signal for generating signals for controlling said noise source; and

means for inserting a signal developed by said noise source into said circuit path of said synthesizer section, thereby to develop voiceless excitation.

6. Apparatus as defined in claim 5, wherein said current magnitude includes low-pass filter means for eliminating selective frequency components from said current signal; and wherein said control signal means includes means responsive to the magnitude of said current for generating a signal representative of the magnitude of said current signal squared, and means responsive to said squared signal and to said area function signal for generating a signal representative of Reynold's number squared.

7. Apparatus as defined in claim 5, further including controllable impedance elements representative of said variable internal impedance being connected in said circuit path of said synthesizer section; and

means responsive to said vocal-tract area function signal and to said current magnitude signal for generat-

7

ing signals for controlling said impedance elements to vary selectively the magnitude of the impedance in said circuit path.

8. Apparatus as defined in claim 7, wherein said impedance control signal generator includes:

means responsive to said area function signal for generating a signal for controlling the impedance value of one of said impedance elements; and
means responsive to the magnitude of said current signal and to said area function signal for generating a

5

10

8

signal for controlling the impedance value of another of said impedance elements.

References Cited

UNITED STATES PATENTS

3,042,748 7/1962 Rosen 179—1

KATHLEEN H. CLAFFY, Primary Examiner

D. W. OLMS, Assistant Examiner