Choice of Base Signals in Speech Signal Analysis*

LADISLAV DOLANSKÝ†, SENIOR MEMBER, IRE

Summary—Fourier series is generally considered to be one of the most basic mathematical tools of the communications engineers; when experimental support is sought for the theoretical conclusions obtained, it has the advantage of having readily available source equipment in the form of sine-wave generators. In many cases, however, due to the characteristics of the signals under study, the analysis into other base functions would result in a reduction of expression complexity and a better insight into the problem. This is demonstrated on a specific example, in which the damped-oscillatory voiced speech sounds are expressed by means of complex-exponential base functions. The method of measuring the pertinent coefficients is given; the nature of the analyzing equipment, which is also used for synthesis, is described briefly, and experimental results, including synthetically obtained approximations of the original signals, are presented.

While speech signals are used to illustrate the method, the latter is applicable to other signals as well.

Introduction

Recent history shows that in speech signal analysis considerable progress has been made without the benefit of precise mathematical expressions for the speech wave forms. More recently, computers are helping to solve certain quantitative aspects of speech analysis. Nevertheless, in order to make better use of the more theoretical tools of mathematics which are available, it would be desirable to have at least an approximate mathematical expression for speech waveforms under study.

The Problem

The problem considered in this paper is to obtain relatively simple expressions for the waveforms of voiced speech sounds in terms of damped oscillatory base functions. While the base functions form an orthonormal set and their shape is fixed for all sounds, the pertinent multipying factors are to be obtained by a simple measurement.

Theoretical Considerations

Damped Oscillatory Base Functions

In view of the quasi-periodic waveform of voiced sounds, the question arises why some standard method of analysis, such as Fourier expansion, should not be used. There are at least three reasons which make the use of undamped sinusoids for base functions undesirable in the case of voiced-sound waveforms. First, while successive pitch periods resemble each other to a considerable degree, the duration of this quasi-periodic function is limited, and thus, Fourier analysis in terms of the fundamental pitch frequency and its harmonics is not strictly applicable (Fig. 1). Second, because of variation of pitch and volume, successive pitch periods seldom have exactly the same waveform. Finally, from the mechanism of generation of voiced sounds, it is known that a pulse-like excitation, originated by the action of the vocal cords, excites the various resonant cavities of the vocal tract and thus starts a combination of decaying oscillatory functions. Thus, an approximation of these decaying functions by ordinary sinusoids does not appear to be too efficient. This is also indicated by the values of the correlation coefficients in certain typical cases. For instance, the correlation coefficient between two damped exponentials $\exp(s_j t)$ and $\exp(s_k t)$ is equal to [3]

$$r_{jk} = \frac{\int_0^\infty \exp(s_j t)\cdot \exp(\bar{s}_k t)\,dt}{\left\{\int_0^\infty \exp[(s_j+\bar{s}_j)t]dt \cdot \int_0^\infty \exp[(s_k+\bar{s}_k)t]dt\right\}^{1/2}} \tag{1}$$

$$= \frac{[(s_j+\bar{s}_j)(s_k+\bar{s}_k)]^{1/2}}{s_j+\bar{s}_k}$$

where

$$s_j = -\alpha_j - j\beta_j = \text{complex frequency},$$

$$\bar{s}_j = -\alpha_j + j\beta_j$$

and

$$\alpha_j \geqq 0.$$

(a)

(b)

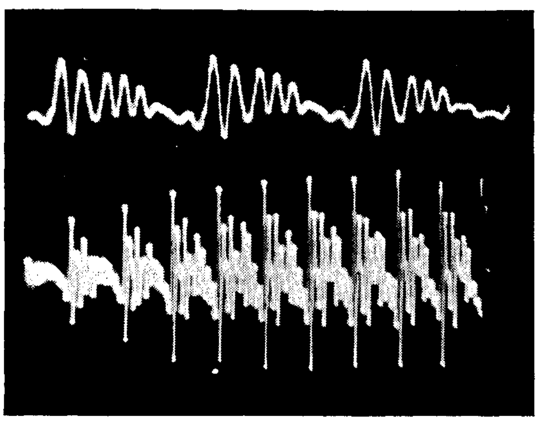

Fig. 1—Waveform of sound *a*. (a) Expanded waveform of the center part. (b) Initial part of the sound.

* Received by the PGA, May 31, 1960. The research reported in this paper has been sponsored by the Geophysics Research Directorate of the Air Force Cambridge Research Center, Air Res. and Dev. Command under Contract No. AF 19(604)-4979.

† Harvard University, Gordon McKay Lab., Cambridge, Mass.

For a sinusoid, $\alpha_j = 0$. Thus, if it is attempted to approximate a damped oscillation by an undamped oscillation, (1) indicates that $r_{jk} = 0$. On the other hand, if a damped oscillation is approximated by another damped oscillation, $\alpha_j > 0$, and $r_{jk} \neq 0$, which indicates that a better approximation may be expected in the latter case.

Using damped oscillatory base functions, the approximation $f_a(t)$ to the given function $f(t)$ becomes

$$f_a(t) = \sum_{i=1}^{n} A_i \exp(s_i t). \tag{2}$$

Other kinds of base functions have been mentioned in the literature. For instance Laguerre functions, which have poles at real values of the complex frequency, have been suggested by Wiener [6]. However, as pointed out by Kautz [4], many problems may be solved more easily with complex poles, because of the greater versatility of damped sinusoids over mere damped real exponentials. This is particularly true in our case because of the nature of the phenomenon considered.

Orthonormality

It has been stated by Huggins [3] that the use of not uncorrelated base functions results in mathematical equations whose solutions are excessively sensitive to slight numerical errors, and that the corresponding instrumentation suffers from the physical counterparts of these same effects. It is therefore most desirable to select base functions which, besides being generated by a process similar to the one which generates the function under study, also form an orthonormal set.

A set of functions which is originally not orthonormal, can be orthonormalized by forming weighted sums of the given functions by adding one of the given functions for each successive base function, $g_i(t)$, and imposing the condition [1]

$$\int_0^\infty g_i(t)\bar{g}_k(t)dt = e_{ik} \tag{3}$$

where

$$e_{ik} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}.$$

In practice, the use of (3) results in considerable algebra. Therefore, Kautz [4] advanced another method of obtaining the orthonormal base functions, based on the complex convolution theorem of the Laplace transform theory. The method uses the fact that orthonormality can also be stated in the complex-frequency domain and is given by the relation

$$\frac{1}{2\pi j}\int_{-j\infty}^{j\infty} G_i(s)\overline{G}_k(-s)ds = e_{ik}, \tag{4}$$

where $G_i(s)$, $G_k(s)$ are the Laplace transforms of the orthonormal base functions $g_i(t)$, $g_k(t)$. For the case of complex poles, the Kautz method leads to (5) and (6) which define the set of orthonormal functions in the complex frequency domain [5].

$$G_{2n-1}(s) = \sqrt{2\alpha_n}\,\frac{s + |\alpha_n + j\beta_n|}{(s+\alpha_n)^2 + \beta_n^2} \cdot \prod_{j=1}^{n-1} \frac{(s-\alpha_j)^2 + \beta_j^2}{(s+\alpha_j)^2 + \beta_j^2} \tag{5}$$

$$G_{2n}(s) = \sqrt{2\alpha_n}\,\frac{s - |\alpha_n + j\beta_n|}{(s+\alpha_n)^2 + \beta_n^2} \cdot \prod_{j=1}^{n-1} \frac{(s-\alpha_j)^2 + \beta_j^2}{(s+\alpha_j)^2 + \beta_j^2}. \tag{6}$$

That is, networks which have transfer functions of (5) and (6) will produce a set of orthonormal time functions at their output terminals when a unit impulse is connected to their input terminals. For simplicity of design, the critical frequencies of (5) and (6) have been chosen according to (7).

$$s_i = -\alpha_i \mp j\beta_i = i(-\alpha_1 \mp j\beta_1), \quad i = 1, 2, 3, 4, 5, 6, 7. \tag{7}$$

Base functions corresponding to $i \geq 8$ have been omitted in view of the fact that with a choice of $\beta_1/2\pi = 400$ cps, the most important frequency range for voiced speech sounds is reasonably well represented by the set of critical frequencies identified by (7). In view of certain preliminary investigations of speech wave forms,[1] the ratio β_i/α_i was chosen to be

$$\beta_i/\alpha_i = 20. \tag{8}$$

Transformation of the Base Functions into the Time Domain

The expressions of (5) and (6) can be transformed into the corresponding time functions by decomposing the multiple product of fractions into a partial fraction expansion and finding inverse Laplace transforms for the individual fractions which can be obtained readily by standard methods. If $\beta_1/2\pi$ is chosen to be equal to 400 cps and $\beta_1/\alpha_1 = 20$, the first two base functions become

$$g_1(t) = 22e^{-125.7t}\sin(2512t + 0.811) \tag{9}$$

$$g_2(t) = 22.95e^{-125.7t}\sin(2512t + 2.381) \tag{10}$$

where t is the time in seconds. The rest of the orthonormal functions contain these functions as well as similar functions having

$$\alpha_i = i\alpha_1 \tag{11}$$

and

$$\beta_i = i\beta_1 \tag{12}$$

[1] Dolanský [2], p. 17.

with an appropriate multiplying factor for each term. These factors follow from the partial fraction expansion of (5) and (6).[2]

Evaluation of Base Function Factors for a Particular Speech Wave

In order to obtain an approximate expression $f_a(t)$ for a particular speech wave $f(t)$,[3] it is necessary to form an appropriate weighted sum of the base functions

$$f(t) \approx f_a(t) = \sum_{i=1}^{n} c_i g_i(t). \tag{13}$$

In other words, it is necessary to find the particular multiplying factor c_i for each base function, which gives the best approximation $f_a(t)$ to $f(t)$. It remains to be shown how these factors can be obtained by measurement.

For orthonormal functions [4], the coefficients of the base functions can be evaluated by means of the relation

$$c_k = \int_0^\infty f_a(t) g_k(t) dt. \tag{14}$$

Now suppose that the signal that we want to approximate is one pitch period of a voiced sound, *i.e.*, let

$$f_a(t) = h(t) \tag{15}$$

where $h(t)$ is the approximate unit-impulse response of the vocal tract, shaped for a particular voiced sound, and can be expressed in terms of the base functions as indicated in (13). Then

$$c_k = \int_0^\infty h(t) g_k(t) dt. \tag{16}$$

Assume that we have a set of filters whose unit impulse responses are the orthonormal, damped oscillatory base functions, and consider the convolution integral

$$v_k(t) = \int_0^\infty e_i(t-u) g_k(u) du \tag{17}$$

where

$v_k(t)$ = output signal of the kth filter at time t,
$g_k(u)$ = unit impulse response of the kth filter,
$e_i(t-u)$ = input signal at time $(t-u)$.

Eq. (17) can be interpreted in physical terms as giving the instantaneous value of the output voltage v_k of the kth filter network which is caused by a continuously varying voltage e_i, connected to the input terminals of the same filter. A partial contribution to the output voltage at time t, caused by a certain value of the input voltage e_i, occurring u seconds earlier, is equal to

$$e_i(t-u) g_k(u) du \tag{18}$$

and if the effects of all past values of e_i are added, the relationship expressed by (17) is obtained.

Now assume that $e_i(t)$ is a particular time function, namely the unit-impulse response of the vocal tract, which has been reversed in time,[4] *i.e.*,

$$e_i(t) = h(-t). \tag{19}$$

Under these conditions,

$$v_k(t) = \int_0^\infty h(u-t) g_k(u) du \tag{20}$$

and

$$v_k(0) = \int_0^\infty h(u) g_k(u) du. \tag{21}$$

It is seen that, except for the variable of integration, (21) is identical to (16), from which the following procedure for measuring the coefficients of the base functions g_k is obtained (see Fig. 2):

1) the transient signal under study (in our case one pitch period of a voiced sound) is recorded;
2) the signal is reversed with respect to time and reproduced;
3) the time-reversed signal is fed into the orthonormal filter set;
4) the voltages at the output terminals of the orthonormal filter at the instant when the time-reversed input signal suddenly ceases[5] are proportional to the desired coefficients.

It should be noticed that, theoretically, $e_i(t)$ must start at $t = -\infty$. However, since $e_i(t)$ is for physical

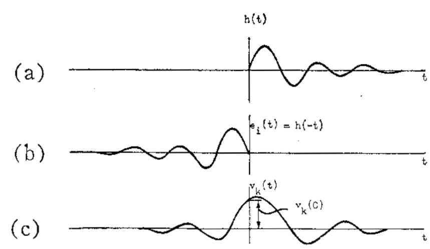

Fig. 2—Transient signal and output of the kth filter. (a) Original transient signal under study (*e.g.*, impulse response of the vocal tract). (b) Same, but reversed in time. (c) Output signal of the kth filter. Note: The instant at which the original transient starts is chosen to be the origin of the time scale t.

[2] A complete list of the fourteen orthonormal base functions thus obtained is given in Dolanský [2], Appendix B, while the pole-zero patterns of the Laplace transforms of (5) and (6) are given in Appendix A.

[3] *E.g.*, for one pitch-period of a particular sound.

[4] *E.g.*, by reversing the direction of a tape recording during playback.

[5] This instant corresponds to the sudden start of the original forward-time transient.

reasons usually negligible for large negative values of t, it is only necessary to choose the beginning of the input at such time t that will make the effect of earlier contributions of $e_i(t)$ to $v_k(0)$ negligible. In practice, the instantaneous voltages $v_k(0)$ were found by measuring the physical size of the instantaneous deflection of one trace of a dual-beam oscilloscope while the other trace of the oscilloscope, displaying $e_i(t)$, was used to identify the instant at which the deflection of the first trace should be read.

Experimental Apparatus

In the preceding section, a four-step procedure for the evaluation of the coefficients of the orthonormal base functions by measurement has been outlined. The apparatus used to perform this procedure will now be described.

Signal Slides

In order to obtain a time-reversed signal of a pitch period of a voiced sound, the microphone signal is first amplified and added to a 500-kc sine wave of suitable magnitude.[6] A Polaroid transparency of such a signal is shown in Fig. 3. This silhouette pattern is then photographically copied onto a high-contrast film. The film copy is used for the reproduction of the original signal in a photoformer.

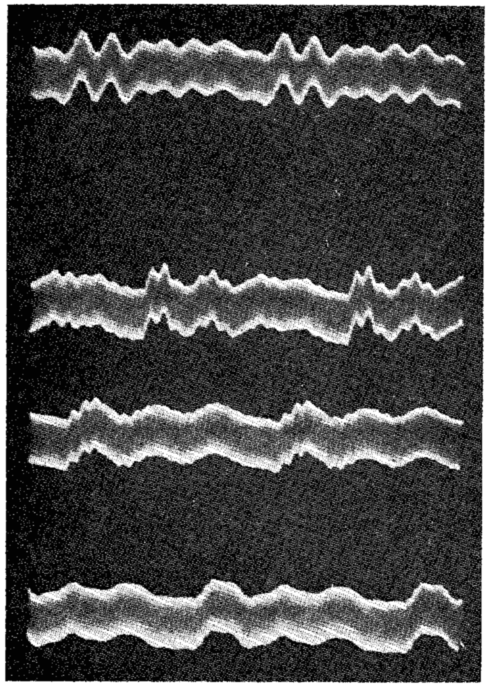

Fig. 3—Photoformer slide.

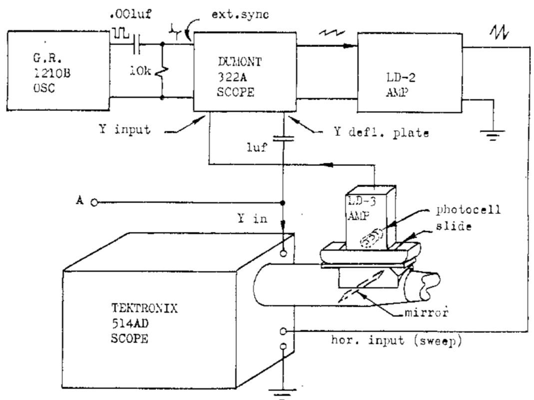

Fig. 4—Photoformer.

Photoformer

The block diagram of the photoformer is shown in Fig. 4. The photocell signal, which is the result of the light generated by the oscilloscope beam, is amplified in the vertical amplifiers of both oscilloscopes and deflects the image[7] of the electron beam of the Tektronix oscilloscope toward the edge of the shadow mask. When the edge of the shadow mask is reached, further deflection of the beam is impossible, since the photocell would no longer obtain the light which causes the deflection. The beam therefore stays in a position which causes its image to fall at the edge of the shadow mask. If the beam is deflected in the horizontal direction at the proper speed, a replica of the original signal appears on the faces of both oscilloscopes, and the signal is also reproduced electrically at A (see Fig. 4). If the pattern is reversed, the reproduced signal appears reversed in time. Incidentally, the same oscilloscope camera is used both for the production of the original slide and for the reproduction of the electrical signal from the slide.

In order to establish to what extent the finite size of the electron-beam light spot in the (photographic) recording process affects the frequency response of the photoformer, the latter was measured in two different ways, keeping the sweep time the same in all cases. In Fig. 5, curve a was obtained by recording an even number of periods of a sinusoid of a certain frequency, but using only one half of the total length of the trace during playback. Thus, for a certain *reproduced* frequency f, the recording frequency must be $2f$. Curve b was obtained by recording the frequency which was desired in reproduction but, before reproduction, the size of the film record was photographically reduced 2:1. During reproduction, the length of the trace, the sweep time and the number of sinusoidal periods per sweep are the same in both cases, resulting in the same *reproduced* frequency, but in case b the finite spot of the recording

[6] The exact frequency used here is not critical, but should always be very much higher than the highest frequency component of the speech signal under study. The high-frequency sine wave is added to the signal in order to obtain a shadow pattern in which the dividing line between the illuminated areas is of the same shape as the original signal, while the area on one side of this dividing line is illuminated because of the presence of the high-frequency sine wave.

[7] The image is formed in the plane of the slide. Alternatively, the slide could be placed directly onto the face of the picture tube, but this results in poorer high-frequency reproduction due to the fact that the slide and the phosphor surface are separated by a glass layer of considerable thickness. Thus, even if the illuminated point is directly behind an opaque point, some light may reach the photocell; this results in a poorer high-frequency response.

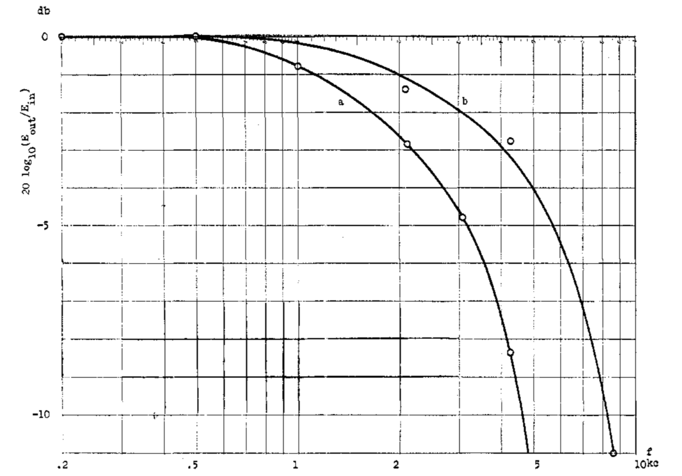

Fig. 5—Photoformer response as a function of the *reproduced* frequency f, (a) when the recorded frequency is divided by two in the playback by sweeping only over one half of the pattern, (b) when the same playback frequency is obtained by reducing a recorded signal photographically by a factor of two.

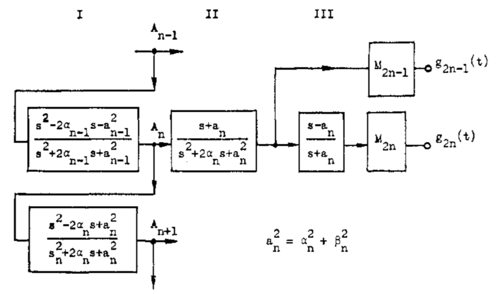

Fig. 6—Orthonormal filter set.

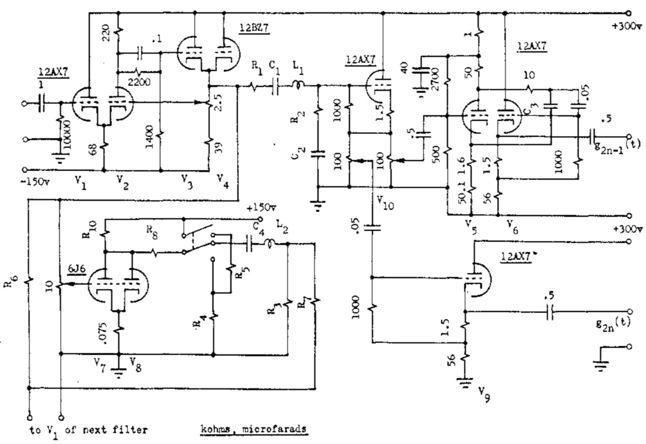

Fig. 7—Circuit diagram for orthonormal filters.

electron beam is photographically reduced 2:1. Therefore, if the recording spot size is the limiting factor, curve *b* should have a cutoff frequency about twice as high as curve *a*. This is indeed confirmed by the curves of Fig. 5.

By a separate measurement it was established that the phosphor persistence[8] of the CRT screen is not a serious limiting factor in our problem. The measurement simply consisted of increasing the sweep rate of the oscilloscope until the reproduced sinusoidal signal decreased to 70.7 per cent of its original value. For the conditions of curve *a* in Fig. 5, the half-power point was reached at a frequency $f_{co} \approx 12$ kc, and for curve *b*, $f_{co} \approx 24$ kc was obtained.[9]

Orthonormal Filters

Based on the theoretical considerations outlined above, a set of filters, capable of generating the first seven functions defined by (5) and the first seven functions defined by (6) has been constructed. According to (7), seven pairs of critical frequencies s_n, $\bar{s}_n$ are involved. Thus, in the scheme developed, all component filters occur seven times, with the response-controlling elements appropriately scaled according to the critical frequency of the particular filter considered.

The construction of the orthonormal filter set used is indicated in Fig. 6. Three basic blocks, I, II and III are used, together with the gain factors M_{2n-1}, M_{2n}, to obtain the desired response. The circuit diagram of the filters I, II, III for one pair of critical frequencies s_n, $\bar{s}_n$ is given in Fig. 7.[10] The computed phase and amplitude response and the experimentally measured points are shown in Figs. 8–11. The amplitude responses of type I and type III filters are not given, since they do not vary with frequency and the deviation of experimental points is only slight. The magnitude response of type II filters deviates somewhat at higher frequencies; this is considered to be unimportant because, in these frequency regions, the relative response of the filters is rather low.

The experimentally obtained impulses are shown in Figs. 12 and 13. It is perhaps somewhat surprising that the sum of the various decaying oscillations gives short trains of sine waves of constant magnitude which suddenly change to a smaller level at regular steps, the spacing between adjacent steps being one period of the lowest-order orthonormal function.

Adding Circuit

While the orthonormal filter set can be used to either generate the orthonormal base functions or to measure the appropriate multipliers of these functions for a particular transient signal under study, some adding circuit capable of forming a weighted sum of the fourteen base functions must be used, if it is desired to reconstruct the original transient signal from the fourteen measured coefficients. The performance of the circuit should be such that the adjustment of one of the coefficients does not affect the remaining coefficients.

The principle of operation of the circuit constructed for this purpose can be explained by means of the basic diagram shown in Fig. 14.

[8] P11 phosphor has been used.

[9] A description of additional experimental equipment used in the photoformer is given in Dolanský [2], ch. V, sec. 1.

[10] More detailed information about the filter circuits can be found in Dolanský [2], ch. V, sec. 3 and Appendix C.

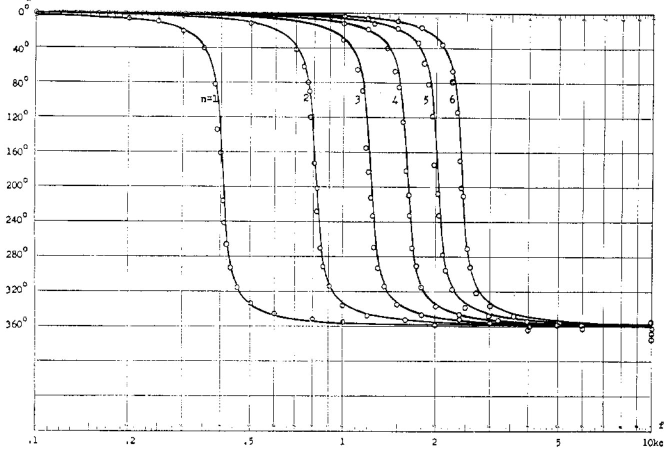

Fig. 8—Phase response of type-I filters.

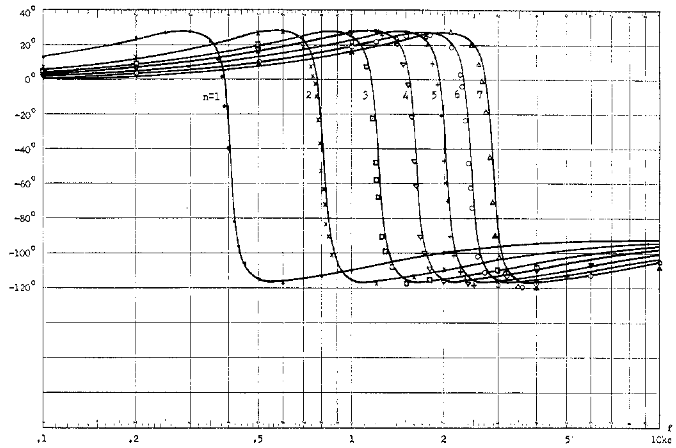

Fig. 9—Phase response of type-II filters.

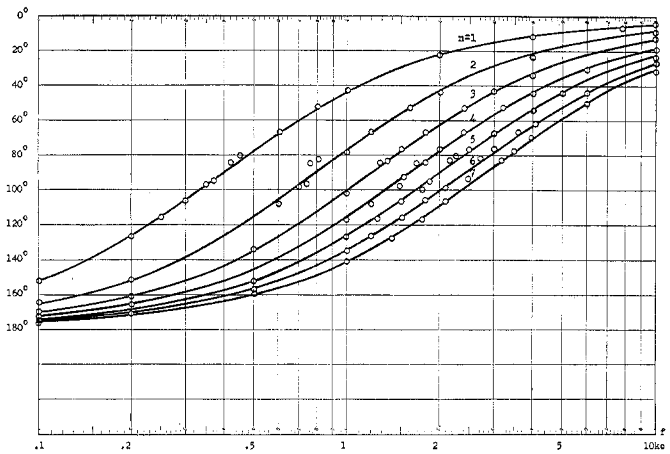

Fig. 10—Phase response of type-III filters.

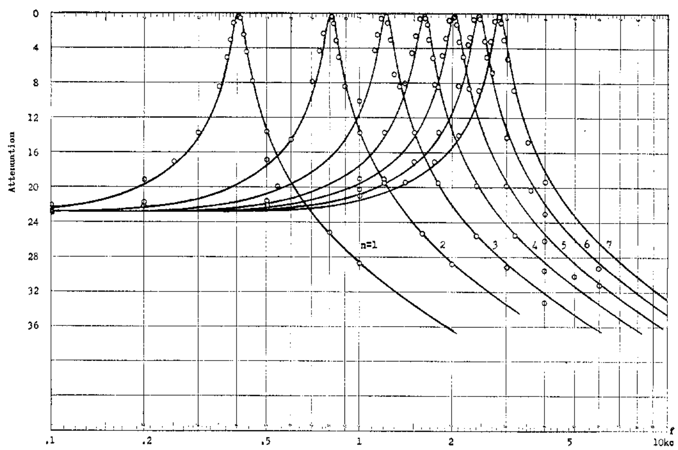

Fig. 11—Magnitude response of type-II filters.

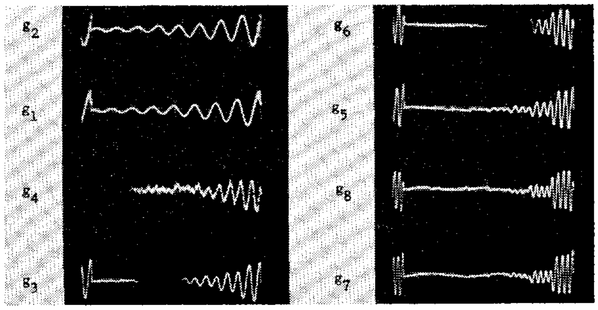

Fig. 12—Unit impulse response of orthonormal filters: $g_1(t)$ to $g_8(t)$. Time increases from right to left. Repetition period 20 msec.

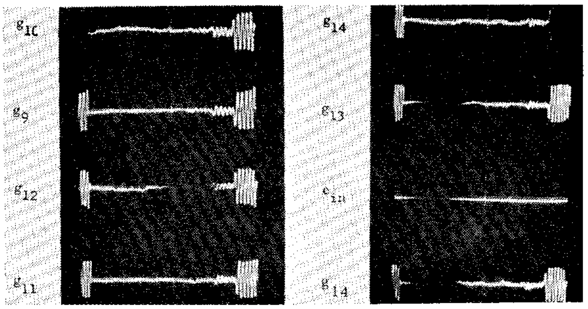

Fig. 13—Unit impulse response of orthonormal filters: $g_9(t)$ to $g_{14}(t)$ and input signal e_{in}. Time increases from right to left. Repetition period 20 msec.

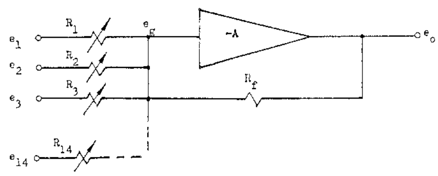

Fig. 14—Basic scheme of adding circuit.

Applying Kirchhoff's current law for the node at the amplifier input, one obtains

$$\sum_{i=1}^{n} \frac{e_i - e_g}{R_i} + \frac{e_o - e_g}{R_f} = 0. \tag{22}$$

Using the relationship between the grid voltage e_g and the output voltage e_o of the amplifier,

$$e_o = - A e_g, \tag{23}$$

(22) can be simplified, for a large gain A, to the relation

$$e_o \approx - R_f \cdot \sum_{i=1}^{n} (e_i/R_i). \tag{24}$$

From (24) it is seen that under these conditions a variation in a particular controlling resistor will affect only the contribution due to the corresponding input voltage e_i, while the effect upon other source voltages will be negligible.

Since each of the constants may have either a positive of a negative sign, two separate adding circuits have been constructed, as shown in Fig. 15. In a particular example, all base functions having multipliers of one sign are added in the same adder. Subsequently, the difference between the positive and negative terms is obtained by means of a difference amplifier. In Fig. 15, V_1 and V_2 represent a cascode amplifier, for which V_3 and V_4 represent a plate load of several megohms. More conventional ways of obtaining the necessary high gain[11] resulted in excessive phase shift which transformed the negative feedback of the circuit into a positive one, thus causing instability.

The degree of the residual mutual dependence of the magnitude-controlling input resistors is indicated by the curves of Fig. 16. Curve *a* represents the output of the adder with the input signal control set for maximum input, while all other inputs are disconnected by means of the switches shown in Fig. 15. Curve *b* indicates the output signal magnitudes obtained for the same input signal as in case *a*, but with all other inputs short-circuited to ground and their control resistors adjusted to their minimum value of 10 kilohms. It is seen that for $0.1 < f < 5$ kc, which includes the range of greatest interest, the signal difference for the two extreme settings is only 0.2 db. Since it is very unlikely that such extreme cases would be encountered, a second resetting of the magnitude controls, after the entire set of them has been adjusted once, should be unnecessary. This has indeed been found to be the case in the practical use of the circuit.

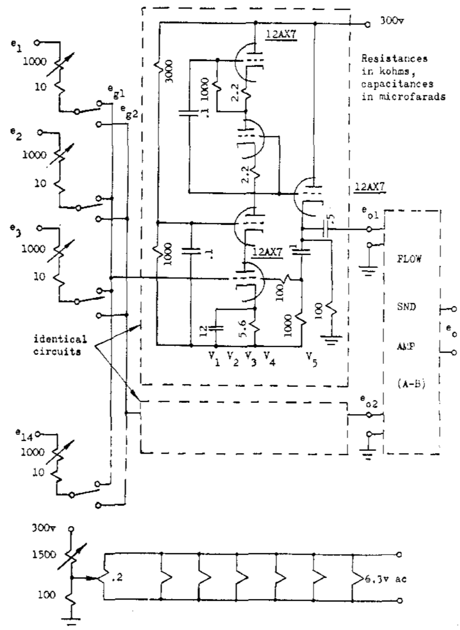

Fig. 15—Adding circuit for orthonormal functions.

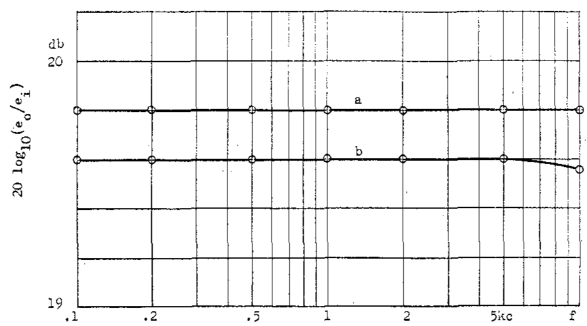

Fig. 16—Response curves of the adding circuit. (a) No loading of grid circuit by other input connections. (b) Maximum loading by other input connections.

Experimental Analysis and Synthesis of Speech Signals

In order to demonstrate the performance of the orthonormal filters on actual speech signals, a limited number of voiced-phoneme signals was analyzed by means of the filter set.

Coefficient Matrices

As indicated previously (13), the transient under study[12] is approximated by a weighted sum of damped-oscillatory base functions, which are fixed by the choice of the particular elements in the filters. The same base functions are used to approximate the waveform of any voiced phoneme and the various phonemes differ only in the values of the coefficients of the individual base functions. Thus, the essential information identifying phoneme resides in these coefficients alone, and can be expressed, for example, as a row matrix

$$M(i) = [c_1, c_2, \cdots c_{14}], \tag{25}$$

which can also be regarded as a vector in a 14-dimensional signal space.

[11] A gain of about 5000 was considered desirable.

[12] *I.e*, one pitch period of a voiced phoneme.

Analysis of Speech Sounds

Following the procedure outlined under theoretical considerations, the coefficients c_i of (25) were measured for 25 phoneme samples. These samples consisted of one pitch period of the sounds *i*, ɛ, *a*, ɔ, *u* as pronounced by five different male speakers. The pitch frequency was held constant at 100 cps. The normalized values of the constants c_i thus obtained were used in the synthesis of speech-sound waveforms described below.

Synthesis of Speech Sounds

In order to show how closely the individual waveforms can be approximated by means of the given base functions, the adding circuit has been used to obtain a weighted sum of these functions. The gain controls for the individual base functions were set to provide multipliers proportional to the constants c_i, obtained in the analysis of the preceding section. The original signal, as obtained from the photoformer, and the signal synthetized from the base functions $g_i(t)$ and from the coefficients c_i are shown in the double oscilloscope traces of Fig. 17. Traces of the same row represent five different sounds pronounced by the same speaker, while all traces of the same column represent the same sound, pronounced by five different speakers. It is seen that for many samples, particularly for all samples of sound *a*, and some samples of sound *i*, ɛ, and *u*, the synthetic waveform is a very good approximation to the original one. In some other cases, especially for the sound ɔ, the agreement is not quite as good. In the case of these poorer approximations, it was usually difficult to determine the exact location of the beginning of the pitch period, which is essential for the measurement of the constants c_i. It is very likely that with a more accurate determination of the exact beginning of the pitch period, better waveform approximation could be obtained in these cases, too. In the case of the phoneme ɔ, the most prominent oscillation seems to be in the vicinity of 600 cps. Since the closest components of the base functions are 400 cps and 800 cps, this sound may be harder to approximate than other sounds which have a prominent component at either 400 or 800 cps. However, as can be seen from trace *a*, AAP (see Fig. 17), a waveform resembling seven cycles in 10 msec can be obtained using the same base functions.

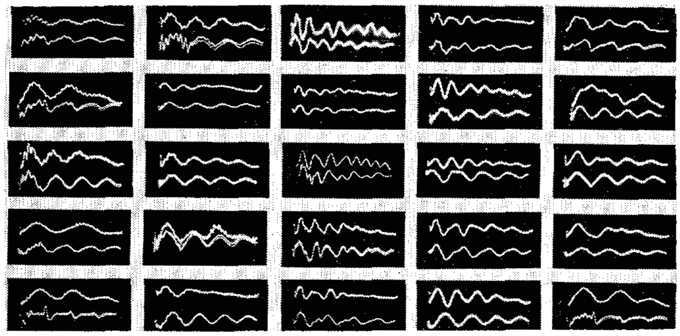

Fig. 17—Original waveforms (upper traces) and reconstructed waveforms (lower traces) of five vowel sounds, spoken by five speakers Speech sounds (left to right): *i*, ɛ, *a*, ɔ, *u*. Voices (top to bottom): DWB, JGB, LD, AAP, AZ.

Conclusions and Limitations

In view of the research reported in this paper, the following conclusions and limitations can be stated:

1) Waveforms of voiced speech sounds can be analyzed and synthetized in terms of a small number of orthonormal, damped-oscillatory base functions, multiplied by appropriate coefficients.
2) These coefficients can be obtained by simple measurement.
3) The result of this analysis-synthesis procedure is a simple mathematical expression in closed form, which can be manipulated by standard mathematical procedures available, in contrast to the original microphone signal or oscillograph display to which these powerful methods cannot be applied.
4) The method of analyzing signals in terms of suitable base functions is not limited to speech waveforms. However, in order to obtain a simple result, the base functions should be carefully selected. Frequently, a study of the process which generates the functions may reveal which functions are more suitable than others in a particular problem under study.
5) Only male voices have been used. Because of the shorter pitch period, more residual signal due to the preceding pitch period may perhaps be expected for female voices, although if the attenuation is approximately constant per cycle, comparable conditions to the case of male voices may be obtained.
6) Turbulent components of voiced sounds are neglected. They would probably be entirely eliminated in a reproduced signal.
7) Although the method is in principle applicable also to waveforms of sounds having only turbulent excitation, in practice it would be difficult to a) establish the exact time of each individual random excitation and b) eliminate interference of residual signals due to recent previous excitations, because of the rapid succession of such excitations.
8) In its present form, the method requires a time reversal of the signal and therefore a recording of the signal which results in a delay. While powerful for basic studies of the characteristics of speech, the method in its present form would not be suitable for a continuous processing of speech in a live transmission link.
9) Frequency normalization with respect to some significant frequency parameter has not been carried out. This may be desirable, particularly if the studies are extended to voices of women and children.

ACKNOWLEDGMENT

The author wishes to express his thanks to Dr. D. W. Batteau, who supervised this research, for his continued support during the entire duration of this work. The scientific advice of Professors P. E. LeCorbeiller and W. P. Raney of Harvard University, Cambridge, Mass., was also very much appreciated.

BIBLIOGRAPHY

[1] R. Courant and D. Hilbert, "Methoden der Mathematischen Physik," Springer Verlag, Berlin, Germany, vol. 1, 1924; Interscience Publishers, New York, N. Y., p. 40, 1943.
[2] L. O. Dolanský, "A Novel Method of Speech-Sound Analysis and Synthesis," Ph.D. dissertation, Harvard University, Cambridge, Mass.; January, 1959.
[3] W. H. Huggins, "Representation and Analysis of Signals, Part 1; The Use of Orthogonalized Exponentials," The Johns Hopkins University, Baltimore, Md., Rept. No. AF 19(604)-1941, AFCRC TR-57-357, ASTIA Doc. No. AD 133741, pp. 7 and 15; September, 1957.
[4] W. H. Kautz, "Network Synthesis for Specified Transient Response," Res. Lab. Electronics, Mass. Inst. Tech., Cambridge, Mass., pp. 6, 16 and 21, Tech. Rept. No. 209; April, 1952.
[5] W. H. Kautz, "Transient synthesis in the time domain," IRE TRANS. ON CIRCUIT THEORY, vol. CT-1, pp. 29–39; September, 1954. See p. 31.
[6] N. Wiener, "Extrapolation, Interpolation and Smoothing of Stationary Time Series," John Wiley and Sons, Inc., New York, N. Y., p. 35; 1950.

The Use of Pole-Zero Concepts in Loudspeaker Feedback Compensation*

WILLIAM H. PIERCE†, STUDENT MEMBER, IRE

Summary—**Pole-zero concepts, with the associated techniques of signal flow graphs and root locus plots, are introduced and used in the analysis and synthesis of integrated loudspeaker-amplifier systems in the lower-frequency region. The case of the infinite baffle system is treated in detail, and formulas are developed for general voltage and current feedback. This is then simplified into a design method using only RC elements, and the compensation of the high-efficiency woofer in an undersized enclosure illustrates the method.**

INTRODUCTION

POLE-ZERO concepts are a very powerful synthesis tool for linear systems, and are so powerful for systems that are of the lumped-constant linear type that the design engineer can synthesize a system from specifications in a straightforward analytical procedure.[1] This paper will show how pole-zero concepts can be applied to loudspeaker systems in the lower frequency range where they may be approximated by lumped, linear elements. In addition to the features that make pole-zero concepts so useful in lumped system synthesis, several topics peculiar to loudspeakers will be discussed. The most fundamental is that in the usual loudspeaker system, there is no electrical voltage or current that is proportional to the sound output, and thus there can be no direct feedback of the acoustic output. Other topics will be analytical procedures for handling voltage and current feedback in terms of pole-zero positions, and a simple way of accounting for variable radiation resistance in calculations of frequency response.

Consider a parallel RLC circuit. The differential equation is

$$i = C\frac{de}{dt} + \frac{1}{R}e + \frac{1}{L}\int_{-\infty}^{t} e(\tau)d\tau.$$

In engineering, this is customarily treated as

$$Y = j\omega C + \frac{1}{R} + \frac{1}{j\omega L}$$

or

$$Y = Cs + \frac{1}{R} + \frac{1}{Ls},$$

where s can be considered to be equal to $j\omega$, or to the operator d/dt, or to the complex frequency in e^{st}, or to

* Received by the PGA, February 3, 1960; revised manuscript received, October 17, 1960.
† 14380 Manuella Ave., Los Altos, Calif.
[1] J. G. Truxal, "Automatic Feedback Control System Synthesis," McGraw-Hill Book Co., Inc., New York, N. Y.; 1955.