The Role of the Pinna in Human Localization

Author(s): D. W. Batteau

Source: *Proceedings of the Royal Society of London. Series B, Biological Sciences*, Aug. 15, 1967, Vol. 168, No. 1011 (Aug. 15, 1967), pp. 158-180

Published by: Royal Society

Stable URL: https://www.jstor.org/stable/75711

The role of the pinna in human localization*

By D. W. Batteau

Professor of Mechanical Engineering, Tufts University, Boston, Mass.

(*Communicated by H. E. Huxley, F.R.S.—Received* 2 *July* 1965
—*Revised* 20 *December* 1966)

The pinna, or external ear, has been shown experimentally to perform an acoustical transformation essential to localization in human hearing. The mathematical form of the transformation is given, the inverse to the transformation is shown to exist, and a theory of localization constructed from the evidence. The theory is extended to reverberation and other facets of human hearing. The basic theorem applied to the study is 'in order to know, the inverse to the transformation of perception must be constructed'.

I. Introduction

The problem of human localization of sound has provided a subject for study which is old and honoured, and the role of two ears in the performance of localization has dominated this study. The pinnae were regarded with interest from time to time, but the major early attitudes regarding the external ear are reflected in a statement by Steinhauser (1878).

'In its essential construction the pinna of the human ear consists of a funnel to collect the sound and a reflector.'

Mach (1874) and Rayleigh (1876) had remarked on the pinna. Mach conjectured that the aspect of the pinna to the sound source might be important, as is indeed the case. It seems appropriate also to mention that it was early observed that the changes in the quality of sound occur with changes in orientation of the source to the hearer (Thompson 1879), and that such quality changes are an unavoidable consequence of transformation. However, most of the early research was directed towards examination of the significance of intensity differences between the two ears as means of localization (Steinhauser 1879; Thompson 1881; Bell 1880). Rayleigh (1907) had also observed that phase difference between the two ears could be important, and as sophistication in science grew, a number of investigators turned their attention to the importance of phase differences in pure tones played separately into the two ears, although the mathematical analyses were available to show the insufficiency of dimension for localization of pure tones by two immobilized ears (Appendix B). Subsequently, the importance of phase received more and more attention, and was shown to be comparatively dominant in localization of sounds by Langmuir and his associates as recently as 1944.

In the attempt to form hypotheses concerning localization, it is useful to consider other aspects of human hearing, such as pitch perception of pure or complex tones, the recognition of speech, and the ability to pay attention selectively to

* This work was performed under the sponsorship of the United States Naval Ordnance Test Station, China Lake, California.

sounds either by locale or character. The means of performing one of these functions must not be incompatible with the performance of other of these functions. Furthermore, if a model can be developed which is consistent with all without invoking multiple hypotheses, such a model is to be preferred. This is merely a paraphrase of Occam's razor. For these reasons, it is important to our consideration that other factors related to the role of the pinna also entered the study of hearing. It was shown by Wallach and his associates (1949) that the first signal to arrive appears to take precedence in a sequence of reverberant sound. In his introduction, Wallach remarks, '... that localization within a reverberant room is both common and useful. The problem of how this is possible remains, however, unsolved'. We consider it significant that the theoretical model, with its well defined mathematical form, presented here for the function of the pinna also provides an adequate explanation of the co-called 'precedence effect'. The mental apparatus for both the role of the pinna in localization and the selection of a sound in a reverberant room is the same except in the lengths of time involved. Indeed, the pinna can be viewed as a fixed reverberator designed to provide different effects for different aspects of the sound source to it.

It is also important to our consideration, since it is consistent with the theory, that studies have been made relative to the perception of phase differences alone as sensations of pitch when binaural interaction is considered with noise (Appendix C). Cramer & Huggins (1948) refer to an unpublished memorandum of Huggins as the earliest thought in this direction. Since pitch sensations by phase difference alone can be produced monaurally with complex wave-forms, as was done by the author in 1958, it is necessary that the mechanisms of hearing be consistent with this observation and significant that the mathematical model derived for monaural localization also provides for the perception of phase changes alone.

In a paper by Kock (1950), a pertinent remark is made in the conclusion: 'The above experiments [referring to his work] suggest that the brain preserves and evaluates time delays [perhaps by the mechanism of delay insertion in one or the other of the nerve paths between the ear and the brain] ...' The mathematical model provided to explain the role of the pinna, the selection from reverberation, the assignment of pitch to noise, the recognition of speech, and the perception of phase, uses in every case only delays and signed additions to provide the requisite functions.

II. Experimentation

The research reported here began in the summer of 1959 with a simple demonstration, performed by W. B. McLean, which implied that the external ears, or pinnae, have a significant role in human localization. McLean (1959) demonstrated that distortion of the pinnae by bending distorts the perception of the locale of the sound.

It is immediately evident that the perception of locale and attention to a sound source require a transformation relative to the arriving wave front, since acoustic information arrives at the ear from the impinging sound front. Since a transformation of the incoming sound front is necessary to localization, it seemed reasonable

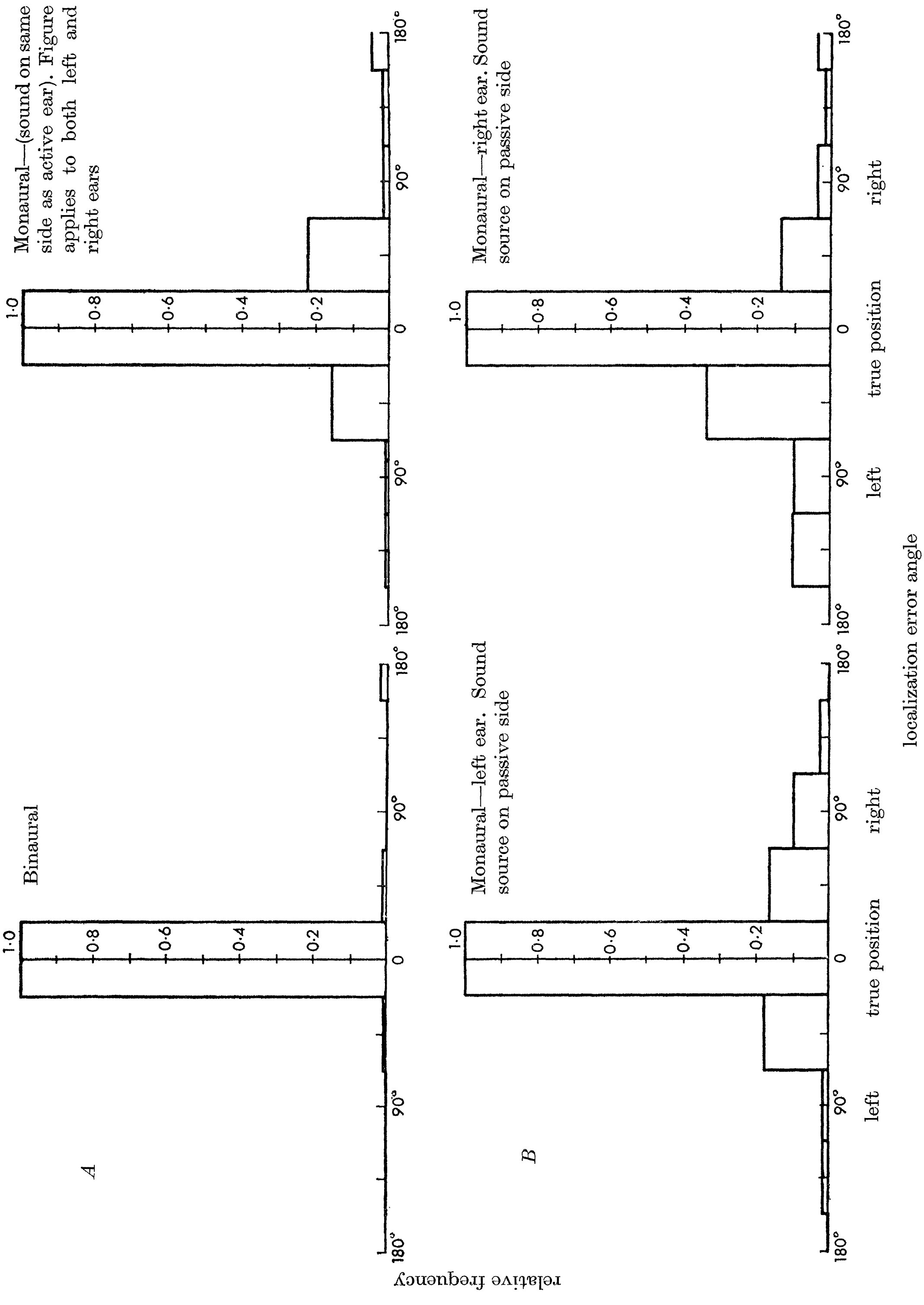

FIGURE 1. Localization histograms for monaural and binaural comparison

to assume that the external ear served as the acoustical device to perform the requisite transformation. Subsequent work (Batteau 1961) demonstrated that it indeed does introduce a significant transformation. Further experimentation showed that the perception of all locales of a sound position (front, back, up, down, left, right, distance) could be provided with a high-fidelity electronic system in which the microphones were inserted in casts of human pinnae (Batteau & Plante 1962). The extraordinary fidelity needed in all aspects of this system, microphones, amplifiers, headphones, acoustic isolation, perhaps has prevented construction of the requisite systems until now. Our most recent work has been done with B & K 4133 microphones, and semi-insertion capacitor headphones of our own construction, which faithfully cover a range from 50 c/s to 40 kc/s.

An early search of the literature brought to light one paper (Angell & Fite 1901) which was concerned with monaural localization, with experimental evidence that it was quite good. With this encouragement, we conducted experiments, using white acoustic noise pulses as a source, with monaural and binaural localization which produced the data presented in figure 1 (reproduced from Batteau & Plante 1962). *Subsequently, it was found that monaural localization by persons totally deaf in one ear is commonplace.*

It is obvious that there are differences in the exact details of individual pinnae, and the results obtained with a group of four colleagues of varying ages from 25 to 38 years, show a variation in performance with the electronic system of pinnae, microphones, and headphones. The polar graph shown in figure 2 presents the results. (The poorest performance was obtained from a man who happened also

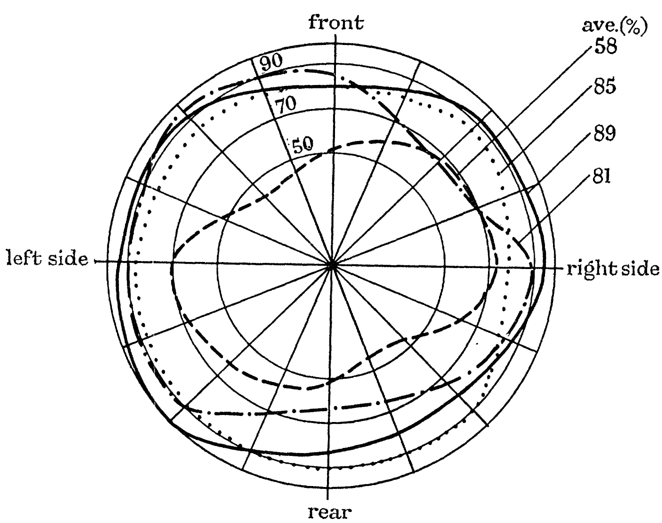

FIGURE 2. Azimuth localization ability with electronic coupling for four subjects.

to be colour blind.) The curves represent averages of 36 trials at each indicated position with a random selection of position. Randomizing was obtained by filling a box with total test figures on individual slips of paper and drawing them one at a time.

Since there are differences in pinna, a test was undertaken to indicate the rate of adaptation to the pinnae used. (It is interesting to observe in figure 3 that two well defined groups were formed by the test subjects.) In order to prevent room acoustics from influencing the performance, the pickup was reoriented during our test; this did not seem to affect the trends development. Each subject was required to make 288 responses, evenly distributed and randomly selected in azimuth. Recent tests in adaptability, conducted by S. Freedman using apparatus developed in our research, also show relative ease of use of a set of foreign pinnae (Freedman & Stampfer 1964). *One may conclude that other people's pinnae can be used.*

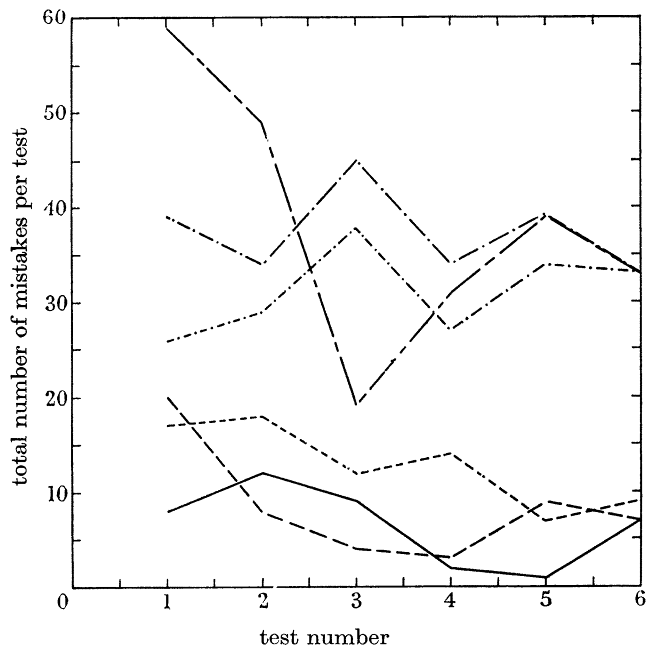

FIGURE 3. The effect of experience with electronic coupling for six subjects. Head reoriented after test no. 3.

In order to obtain information concerning the transformation performed by the pinna, an ear cast and microphone were used to provide oscillograms of the response to a pulse. A sample of this work is shown in figure 4. The form of the pulse as recorded by a bare microphone is shown in 4*a* and the oscillogram taken through the pinna is shown in 4*b*. The appearance of the oscillograms is much more suggestive of sets of discrete delays than of a smooth continuum, although consideration for both is provided in the theory.

By modelling the ear five times normal size and placing a microphone in it, R. Plante and W. Lyle were able to obtain a measurement for delays which vary with azimuth and also for those which vary with elevation angle. The results of these measurements, scaled to times equivalent to the normal sized ear, are presented in figure 5 (reproduced from Batteau & Plante 1962).

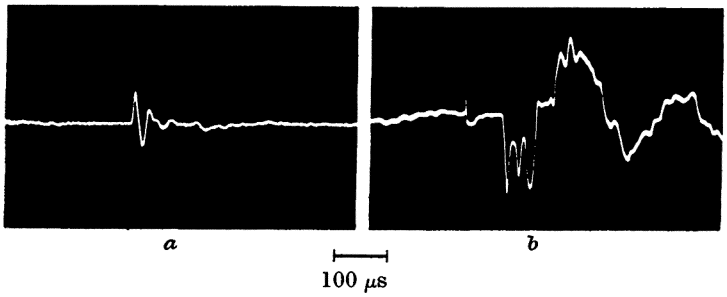

FIGURE 4. (*a*) The test pulse. (*b*) The transformation of the test pulse by the pinna to the side.

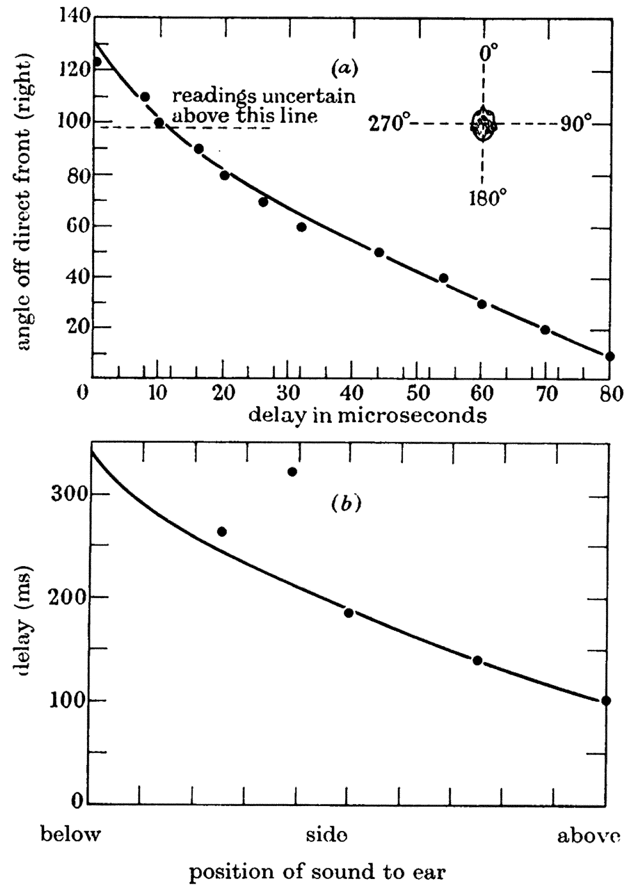

FIGURE 5. Time delay against orientation. (*a*) Azimuth, (*b*) elevation.

III. DEVELOPMENT OF THEORY

When it is considered that attention is one of the prime functions of localization, a new factor is introduced into the problem. When signals are presented separately to the two ears, it is conceivable that correlation can be performed on a selected signal if there is a distinct amount of time difference for it between the two ears, compared with other time differences for other signals. It may also be possible to select one ear or the other with which to listen by choosing intensity differences. However, neither anywhere in the surveyed literature nor in prior localization theory is it possible to account for the relatively acute *localization of attention* by the healthy young, nor the so-called 'cocktail party effect' on the basis of two ears alone. The mathematical model based on computation of a structural transformation by multiple reflexions and delays by the pinnae does, however, provide for both attention and localization. In fact, the locality is determined by the structure of the transformation necessary to pay attention to a sound at a given place.

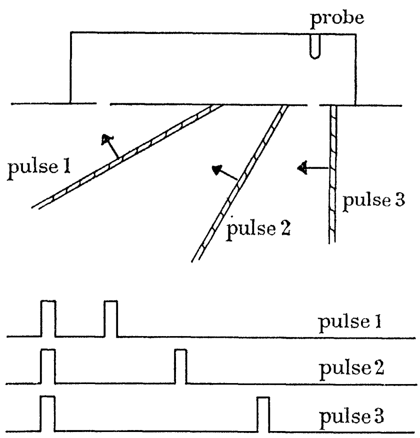

FIGURE 6. A directional coupler in a plane

The working hypothesis regarding the pinna was very simple. It assumed that the various paths through the pinna, which the sound must travel to reach the ear canal, involved multiple delays because of path length differences. The delays introduced by the pinna were easily seen to depend on the orientation of the sound source with respect to the pinna. In order to examine the significance of delay and orientation, a simple diagram of a system reduced to one delay, which is called a directional coupler, is shown in figure 6. The three pulses shown are assumed to come from different directions. The arrows indicate the directions of travel of the sound fronts to which they are attached. The sound enters each of the two ports at different times, and thence to the microphone with the time separations shown in the lower diagram.

In the simple model shown the pulse separation as a function of angle is given by the following equation:

$$\tau = s/v\ (1+\cos A),$$

where τ = delay between first and second arrivals at the probe; s = distance between holes in the coupler; v = velocity of sound in the medium; A = angle of normal to wave from t with line between holes: range 0° to 180°.

If a sound picked up by the microphone in the directional coupler can be restored to its original form, then the transformation providing that restoration is the inverse of the transformation of the coupler, and the direction of the wave front can be found from the inverse transformation. It is possible to construct inverse transformations in several ways. The first of such inverses to be constructed is as follows.

Let $f(t)$ = original sound at source,

$f(t)+af(t-\tau)$ = sound at ear drum due to $f(t)$ coming through the simple delay system,

τ = time delay in reflexion path,

a = attenuation factor at path reflexion.

Construct the series

$$\begin{aligned} h(t) &= \sum_{n=0}^{M}(-1)^n a^n f(t-n\tau)+\sum_{n=0}^{M}(-1)^n a^{n+1}f(r-\tau-n\tau) \\ &= f(t)+\sum_{n=0}^{M}(-1)^n a^n f(t-n\tau)+\sum_{n=1}^{M}(-1)^{n-1} a^n f(t-n\tau) \\ &= f(t)+\sum_{n=1}^{M}(-1)^n a^n f(t-n\tau)-\sum_{n=1}^{M}(-1)^n a^n f(t-n\tau)+(-1)^M a^{M+1}f(t-\tau-M\tau) \\ &= f(t)+(-1)^M a^{M+1}f(t-\tau-M\tau) \end{aligned}$$

$$\lim_{M\to\infty} h(t) = f(t) \quad (a < 1). \tag{1}$$

A second method of computing the inverse includes the provocative series of octaved delays, i.e. $\tau, 2\tau, 4\tau, \ldots, 2^n\tau$, which suggest a basis for octave identity in music.

Let $P_1(t) = f(t)+af(t-\tau)$ be the signal after delay and addition. Then construct a sequence:

$$P_n(t) = P_{n-1}(t)+a^{2n-2}(-1)^{2n-2}P_{n-1}(t-2^{n-2}\tau) \quad (2 \leqslant n \leqslant \infty)$$

and by examination, we find in consequence that

$$P_n(t) = f(t)-a^{2n-1}f(t-2^{n-1}\tau).$$

Using the rule of construction and P_n, we find

$$\begin{aligned} P_{n+1}(t) &= f(t)-a^{2n-1}f(t-2^{n-1}\tau) \\ &\quad +a^{2n-1}(-1)^{2n-1}f(t-2^{n-1}\tau)-a^{2n-1}a^{2n-1}(1)^{2n-1}f(t-2^{n-1}\tau-2^{n-1}\tau) \\ &= f(t)-a^{2n-1}f(t-2^{n-1}\tau)+a^{2n-1}f(t-2^{n-1}\tau)-a^{2n-1}a^{2n-1}f(t-2^{n-1}\tau-2^{n-1}\tau) \end{aligned}$$

and $$(a^{2n-1})^2 = a^{2n}$$

and $$2^{n-1}\tau+2^{n-1}\tau = 2^n\tau;$$

so $$P_{n+1}(t) = f(t)-a^{2n}f(t-2^n\tau). \tag{2}$$

By finite induction, starting with P_2, the expression of P_n, P_{n+1} in terms of $f(t)$, etc., is correct. Finally, $\lim_{n\to\infty} P_n = f(t) \quad (a < 1)$.

A third way to construct the inverse to the transformation makes use of negative feedback and is best expressed in the notation of the Laplace transform (see Appendix A). In this case, a delay is indicated by the equation

$$e^{-s\tau} = \text{a delay of } \tau.$$

And the signal at the microphone of the directional coupler is given in equation

$$H(s) = \text{signal at microphone}$$
$$= P(s)\,(1+a\mathrm{e}^{-s\tau}),$$

where $P(s)$ = incoming signal; a = attentuation in the delay path and $a < 1$.

The inverse to the transformation is given in the following equations:

$$A(s) = \text{inverse to coupler transform,}$$
$$A(s) = (1+a\mathrm{e}^{-s\tau})^{-1}$$
$$= \sum_{n=0}^{\infty} (-1)^n\, a^n\, \mathrm{e}^{-sn\tau}$$

and is the Laplace form of the first inverse. The feedback circuit of figure 7 is described by the following equations

$$E(s)\,(1+a\mathrm{e}^{-s\tau}) = H(s). \tag{3}$$

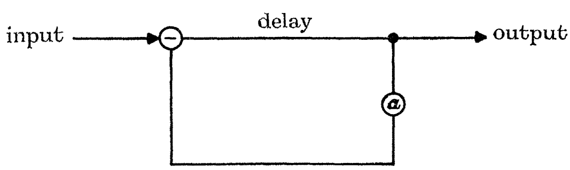

FIGURE 7. A feedback delay computation.

Thus if $H(s) = P(s)\,(1+a\mathrm{e}^{-s\tau}),$

then $E(s) = P(s).$

It should also be pointed out that any means of performing an inverse transformation will meet the logical requirement; the ones presented have seemed most consistent with the observed structure of the nervous system. In particular, however, the power density spectrum of a frequency analysis does not meet the requirement for a complete transformation although the complex Fourier transform does.

In the one-dimensional model of the coupler, a means of measuring the time between first and second arrivals, by constructing an inverse transform or other means provides the direction of the incident wave. If a higher dimensional set of such measurements is desired, a spanning set of holes can be used, with domain definition for each coordinate. For example, three holes in a plane provide a spanning set for two directional measurements or a two-dimensional space. These may be arranged in a rectangular coordinate array, with a fixed delay between one of them and the probe to provide domain separation, as given in the following pair of equations:

$$\tau_1 = \frac{s_1}{v}(1+\cos A_1), \tag{4}$$

$$\tau_2 = \frac{ks_1}{v} + \frac{s_2}{v}(1+\cos A_2), \tag{5}$$

where s_1 = spacing between reference and azimuth holes; s_2 = spacing between reference and altitude holes; A_1 = azimuth angle; A_2 = elevation angle and $k \geqslant 2$.

In these equations, the coordinate system is one which refers to the angle formed between the normal to the incident wave front and the line drawn between the reference hole and the pertinent coordinate hole. A sketch of a three-hole coupler is shown in figure 8.

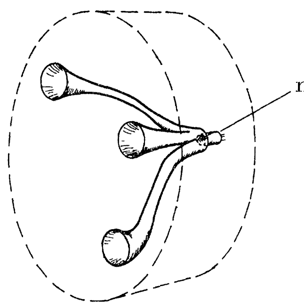

FIGURE 8. A three-hole coupler for azimuth and elevation.

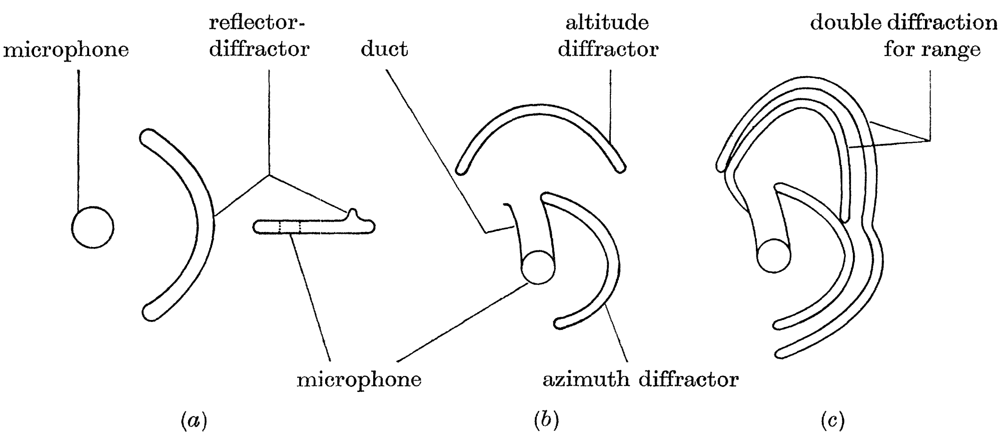

FIGURE 9. Evolution of a diffraction system providing localizing delays in (*a*) azimuth, (*b*) altitude and (*c*) range.

When the consideration of range or curvature sensing is added to this system, it is possible to conceive of displaced systems of coordinates providing the sensing of curvature by relative delays. An evolution of a system is shown in the sketches of figure 9.

While the resultant geometries are provocatively like the structures of the pinna, these should be taken only as suggestive and not as defined modelling. While the theoretical base appears to be sound, many different geometries are possible providing delay transformations, as may be seen by examination of the base of the ear of dog, cat, mouse, bat, and other mammals.

In extending the simple one delay system to the role of the pinna, we must consider that a spanning set of delays for localization in three space has a minimum of four linearly independent components, and this dimensionality must be preserved by the pinna, requiring a minimum of four independent paths for the sound to reach the ear drum. It is convenient to continue the Laplace transform notation. The signal in this representation is

$$H(s) = [1+a_1 e^{-s\tau_1}+a_2 e^{-s\tau_2}+a_3 e^{-s\tau_3}]\, F(s), \tag{6}$$

where $H(s)$ = sound heard at ear drum and $F(s)$ = source sound for a spanning set. However, we might generalize

$$H(s) = \sum_{k=0}^{N} a_k e^{-s\tau_k}[F(s)] = T(s)\, F(s). \tag{7}$$

The inverse for this in transform representation becomes

$$T^{-1}(s) = \sum_{n=0}^{\infty} (-1)^n \left[\sum_{k=1}^{N} a_k e^{-s\tau_k}\right]^n. \tag{8}$$

The time domain realization of $T^{-1}(s)$ is then a series of delays with particular attenuation factors. Obviously an approximation

$$T^{-1}(s) \doteqdot \sum_{n=0}^{L} (-1)^n \left[\sum_{k=1}^{N} a^k e^{-s\tau_k}\right]^n$$

converges to some tolerance when the reflexion factors are less than unity.

Having by construction established the existence of inverses, and hence the production of localizing information, it was possible to consider other mathematical treatments in order to explore the function of attention. Although the spanning set of signals for localizing attention is four, we now wish to consider a general proposition.

If a finite set of delayed signals is considered, having τ_M = maximum delay,

$$R(s) = F(s) \sum_{n=0}^{M} a_n e^{-s\tau_n} \tag{9}$$

and the transformation $T(s)$ constructed

$$T(s) = \sum_{n=0}^{M} a_n \exp\{-s(\tau_M-\tau_n)\} \tag{10}$$

the resultant signal $P(s)$ is given

$$\begin{aligned} P(s) &= R(s)\, T(s) \\ &= F(s) \sum_{j=0}^{M} a_j \exp(-s\tau_j) \sum_{k=0}^{M} a_k \exp\{-s(\tau_M-\tau_k)\} \\ &= F(s) e^{-s\tau_M} \sum_{j=0}^{M} a_j e^{-s\tau_j} \sum_{k=0}^{M} a_k e^{+s\tau_k}. \end{aligned} \tag{11}$$

The result is the original signal delayed by τ_M and modified by the product of sums. The product of sums can be represented as follows:

$$\sum_{j=0}^{M} a_j e^{-s\tau_j} \sum_{k=0}^{M} a_k e^{+s\tau_k} = \sum_{n=0}^{M} a_n^2 + \text{cross-terms} \tag{12}$$

$$\text{cross-terms} = \sum_{k=0}^{M} \sum_{j=0}^{M} a_k a_j e^{-s\tau_k} e^{+s\tau_j} \quad (j \neq k).$$

In this product, the cross-term result for $j \neq k$ can be given as follows:

$$\sum_x a_x \, (e^{-s\tau_x} + e^{+s\tau_x}), \qquad (13)$$

where $\tau_x = \tau_k - \tau_j \quad (j \neq k)$

and $a_x = a_{kj} = a_{jk}$.

This form indicates a symmetrical shift about the maximum delay in the general expression, and in the frequency domain indicates no phase shift of components but an amplitude change between $+a_j a_k$ and $-a_j a_k$. Thus such terms can be called 'coloration'. The general expression then becomes

$$P(s) = F(s) \, e^{-s\tau_M} \left[\sum_{n=0}^{M} a_n^2 + \text{coloration} \right]. \qquad (14)$$

The perceived signal would then be increased in amplitude by

$$\sum_{n=0}^{M} a_n^2$$

and coloured by terms of the form $a_j a_k \, (e^{-sx} + e^{+sx})$, none of which could alter the result by more than $\pm a_j a_k$.

For example, let

$$a_0 = 1, \quad a_1 = 0{\cdot}9, \quad a_2 = 0{\cdot}8, \quad a_3 = 0{\cdot}7, \quad a_4 = 0{\cdot}6.$$

The increase in amplitude, I, is then $I = 3{\cdot}30$ and the maximum coloration, C_M

$$C_M = a_0 a_1 = 0{\cdot}9.$$

The subjective result of this method of attention would be: (*a*) increased loudness; (*b*) perceived with delay τ_M; (*c*) coloured somewhat in spectrum.

If we wish to examine the limit of attentive selection provided by this function, we can assume that all reflexion coefficients are unity, and write the expression for the resultant power comparing two original signals of equal magnitude but at different locations. In this case the two signals $H_1(s)$, $H_2(s)$ can be written as follows:

$$H_1(s) = P_1(s) \sum_{n=1}^{N} a_n e^{-s\tau_n}, \qquad (15)$$

$$H_2(s) = P_2(s) \sum_{k=1}^{K} a_k e^{-s\tau_k}. \qquad (16)$$

The attention transform, $A_1(s)$ is formed for $H_1(s)$, and it is assumed that the two transformed primitive sounds P_1, P_2 are added to give a resultant $R(s)$.

$$A_1(s) = \sum_{n=1}^{N} a_n \exp\{-s(\tau_M - \tau_n)\}, \qquad (17)$$

$$R(s) = H_1(s) + H_2(s). \qquad (18)$$

Then the attention function is applied to $R(s)$

$$R(s) \, A_1(s) = e^{-s\tau_M} \left[\sum_{n=1}^{N} a_n^2 + \text{cross-terms} \right] + e^{-s\tau_M} \left[\sum_{n=1}^{N} a_n e^{+\tau_n} \sum_{k=1}^{K} a_k e^{-\tau_k} \right]. \qquad (19)$$

If all of the coefficients are taken as unity, the power in the first term of the expression, W_1, is

$$W_1 = N^2+N^2-N, \tag{20}$$

and in the second term is W_2

$$W_2 = NK;$$

and the ratio of powers resultant is given by

$$\frac{W_1}{W_2} = \frac{2N}{K} - \frac{1}{K}, \tag{21}$$

and if the minimum spanning set is postulated

$$\frac{W_1}{W_2} = 1\tfrac{3}{4}$$

or slightly less than 3 dB of selection by use of the pinna. When two ears are involved, the same kind of computation shows for binaural powers B_1, B_2

$$\frac{B_1}{B_2} = 3-\frac{1}{N} = 2\tfrac{3}{4} \tag{22}$$

or slightly less than 4·5 dB. Our initial measurements have shown a selection of 3·8 dB for speech intelligibility relative to a 60° separated white noise source. More experimentation needs to be done in this area.

In order to complete a set of characterizations by delay, we may consider a continuous reflecting and delaying system. The pinna is in fact such a system although the oscillograms show prominent and well-defined delays. The expression for the complete transformation provided by the pinna may be given by summing over all the delays, giving

$$\int_0^{T_{\max}} a(\tau)\,\mathrm{e}^{-s\tau}\,\mathrm{d}\tau = \int_0^{\infty} a(\tau)\,\mathrm{e}^{-s\tau}\,\mathrm{d}\tau = A(s), \tag{23}$$

where $$a(\tau) = 0 \quad \text{and} \quad \tau > \tau_{\max}.$$

If the incoming signal is presented in its Laplace transform notation, $F(s)$, the sound reaching the ear drum becomes

$$F(s)\,A(s) = H(s),$$

where $$H(s) = \text{sound transform at ear drum.}$$

The inverse function to be performed by mental computation provides

$$F(s) = \frac{H(s)}{A(s)}. \tag{24}$$

IV. Applicability to nervous system

In order to examine the construction of inverses, or attention functions, it is again convenient to use a single fixed delay system. If it is assumed that a characteristic delay τ has been inserted by the pinna, a possible nerve system constructable at the basilar membrane, based on the first mathematical model is as shown in figure 10. This model assumes that the cochlea provides a delay line, with the basilar membrane carrying the delay connexions. It should be noticed that the attenuation constants a^n behave as if there were constant attenuation per unit delay, providing a time-density equivalence in the network.

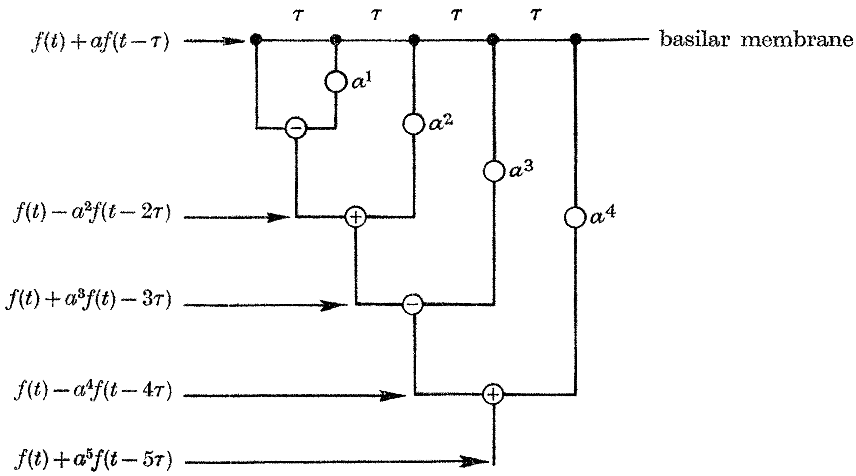

FIGURE 10. Theoretical nerve end and synaptic distribution at the basilar membrane as a computational system.

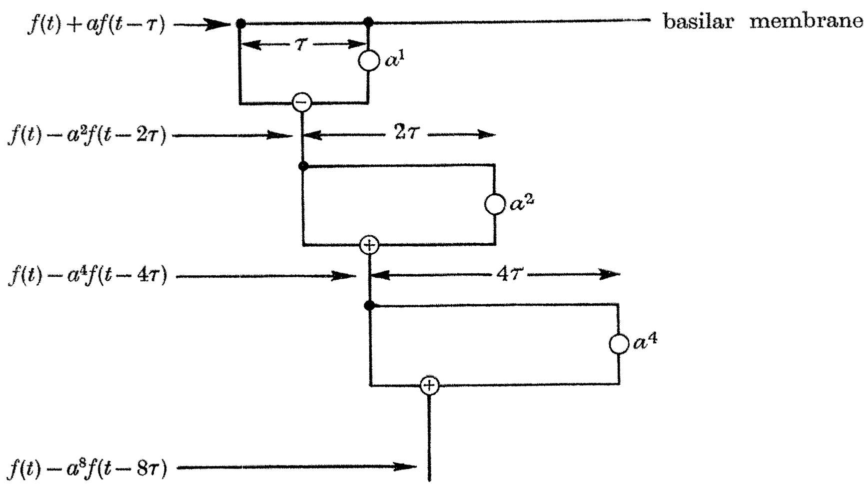

FIGURE 11. Computation of inverse to pinna transform by octaves.

If we use the second mathematical model, it is necessary to introduce nerve delays, and computation may be performed en route from the nerve endings at the basilar membrane to the auditory centre. Figure 11 shows a model of this network.

One of the provocative outcomes of this research has been the construction of mathematical models which appear to be realizable in the human nervous system by rather simple means. If hypotheses are to be constructed regarding the arrangement of the auditory nervous system based on the provided mathematical model, it seems desirable to state plainly the position with respect to the hypothesis of von Bekesey regarding pitch perception as a consequence of mechanical resonances of the basilar membrane, as well as the hypothesis here assumed.

(1) Mechanical resonance of the basilar membrane has no significant role in perception.

(2) The construction of networks by delays and signed additions provides all processing for attention, localization, and recognition.

With the general hypothesis it may further be hypothesized that localization functions can be constructed directly at the basilar membrane. However, attention in reverberation and recognition of speech and music appear to require later structures to provide the necessary longer delay. Pitch perception by correlation of the form described requires a maximum delay of one half cycle for a minimum correlating network. Perhaps the low frequency response of hearing is indicative of the actual correlation lengths involved.

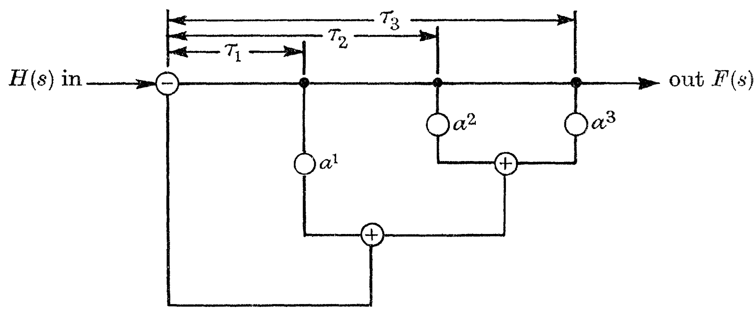

FIGURE 12. A feedback computation inverting three delays.

When the feedback model of inverse is considered, the requisite network for a pinna inverse is constructed as shown in figure 12. If the primitive sound $P(s)$ is transformed by $T(s)$ into $H(s)$

$$H(s) = P(s)\,T(s), \tag{25}$$

$$T(s) = \sum_{n=0}^{3} a_n e^{-s\tau_n} \tag{26}$$

it is convenient to normalize by postulating

$$a_0 = 1, \quad \tau_0 = 0.$$

The feedback network provides the transformation

$$E(s)\,(1 + \sum_{n=1}^{3} a_n e^{-s\tau_n}) = H(s), \tag{27}$$

where $$E(s) = P(s).$$

V. Extensions of theory

In considering the significance of the pinna, attention, and computation, it seemed useful to extend the model of inverse computation to include other domains. Six domains are identified tentatively:

pinna	2 μs to 300 μs
interaural	0 μs to 800 μs
speech	0·6 ms to 2·6 ms
shape	2·6 ms to 10 ms
reverberation	10 ms to 40 ms
memory	0 to ∞

These domains certainly may overlap, but the pinna domain can be synthesized by pulse techniques and sensations of motion obtained, so that it may be called a 'where' domain. The same is true and well known for the interaural domain which may also be called 'where'. The speech and shape domains include the lengths characteristic of the vocal tract and musical instruments and may be called 'what'. The reverberation domain, of course, includes the system of reflexions from walls and objects in the environment and may be called 'place'. The memory domain includes all delays as well as factors of long delay used in personal and environmental recognition. A paraphrase could be made regarding the domains as the recognition of 'something' (what), 'somewhere' (localization) in 'someplace' (environmental) as having some significance (memory).

When 'where' and 'what' are considered, the significance of the median plane becomes apparent. Sound sources lying in the median plane have identical pinnal transforms on the left and right, so that no separation is provided between 'where' and 'what'. Away from the median plane, each ear transforms differently, so that 'where' is different for each ear, but 'what' is the same, providing distinguishability.

The characterization of reverberation, in particular, has the identical form of the expressions given for the pinna, so that mental inversion of the resultant transform can provide attention dependent upon location within the environment, the delay characteristics being unique from point to point. Simple tests have shown two significant facts: (i) in monaural subjective tests, apparent reverberation is reduced by a pinna compared with no pinna, (ii) subjective loudness is raised in a reverberant environment for the same source acoustic power (McNay *et al.* 1961). The same mathematical models apply to these processes as apply to the pinna transformation.

A theoretical process including the role of the pinna in normal hearing can be stated as follows:

1. The incoming sound is characteristic of its source, transformed by the environment.

2. The incoming sound is transformed by the pinna and each direction of arrival has a characteristic.

3. The transformations pertinent to the source to which attention is to be paid are inverted for all directions of arrival, giving a set of environmental transformations.

4. Each environmental transformation for the sound to which attention is paid is inverted and the resultant set of elements combined.

5. The same process is accomplished by each channel (right and left ears) and the results combined.

6. The resultant signal is inverted to the most acute stimulus and interpreted or recognized by the requisite transformation used to accomplish the inverse.

When combination is mentioned, it connotes identity in signal except for time shift, so that simple delay and addition are sufficient. Since the inverse of the environmental transforms provides a combining set, a precedence effect is a simple consequence.

The neural networks necessary to accomplish this process require delays, attenuations, and signed additions. Is its apparent that tone recognition, both pure tones and coloured noise, can be perceived in this manner, and it is theorized that no mechanical resonances of the basilar membrane or other acoustic structures are significant in the process.

Summary

It may be said in summary that, in theory, the role of the pinna in localization is to introduce, by means of delay paths, a transformation of the incoming signal which is mentally inverted to provide attention, and that the inverse transform required defines the location of the sound source. It may be further shown that relatively simple systems of delays, attenuations, and signed additions may be used to construct the inverse transformations, and that these could easily be realized in the nervous system. It may be further theorized that the same method of constructing inverse transformations can apply to monaural and binaural localization, sound recognition, and the utilization of reverberation.

Appendix A

In the examination of the effects of delayed signals, the Laplace transform provides a convenient tool. It is defined as follows:

$$L f(t) \overset{\Delta}{=} \text{Laplace transform of } f(t) \tag{A. 1}$$

$$L f(t) \overset{\Delta}{=} \int_0^\infty e^{-st} f(t)\, dt, \tag{A. 2}$$

$$S \overset{\Delta}{=} \text{a complex variable}, \tag{A. 3}$$

$$(\overset{\Delta}{=} \text{ means 'defined as'}), \tag{A. 4}$$

$$f(t) \overset{\Delta}{=} 0 \quad \text{if } t < 0. \tag{A. 5}$$

The transform may also be written using a unit function to set the requirement (A. 5).

$$U(t) = 0 \quad (t < 0), \tag{A. 6}$$

$$U(t) = 1 \quad (t \geqslant 0), \tag{A. 7}$$

$$L\,f(t) = \int_0^\infty e^{-st} f(t)\, U(t)\, dt. \tag{A. 8}$$

Then a delay in time of the function can be written as (A. 9)

$$f(t-\tau)\, U(t-\tau) \tag{A. 9}$$

and the Laplace transform of the delayed signal written as (A. 10)

$$L\,f(t-\tau) = \int_0^\infty e^{-st} f(t-\tau)\, U(t-\tau)\, dt. \tag{A. 10}$$

Because of the nature of $U(t)$ (A. 10) can be written as (A. 11)

$$L\,f(t-\tau) = \int_\tau^\infty e^{-st} f(t-\tau)\, U(t-\tau)\, dt \tag{A. 11}$$

and by a change in arguments

$$L\,f(t-\tau) = \int_0^\infty e^{-s(t+\tau)} f(t)\, U(t)\, dt. \tag{A. 12}$$

From (A. 12) it follows that

$$L\,f(t-\tau) = e^{-s\tau} \int_0^\infty e^{-st} f(t)\, U(t)\, dt \tag{A. 13}$$

$$= e^{-s\tau}\, Lf(t). \tag{A. 14}$$

Thus multiplication by $e^{-s\tau}$ of the Laplace transform of a time function is equivalent to delaying that function by τ. The algebraic simplification consequent accounts for our use of this form.

Appendix B. Mathematical aspects of geometry

There are only two variables for a pure tone, phase and amplitude. In any channel in which linear summation is performed, the result of transformation of a pure tone can alter only those parameters. In two channels, then, only a comparison of relative amplitudes and relative phases can be made, thus a space of two dimensions is the maximum constructable from pure tones in two channels and is insufficient to describe a three-dimensional space. Such a system would produce lines of constant value, or ambiguity, in the surroundings.

In the treatment of signals as vectors, as initiated by David Hilbert (Courant & Hilbert 1953) and greatly applied by Norbert Wiener (1949) a pure tone is a single vector in a complex function space. As such, it is insufficient to define a three-dimensional volume. If the argument, or relative phase angle, of such a vector is considered alone, four pure tones are a minimum to provide a basis for a three-space in a single channel. One tone serves as referent and the other three as referred to form a coordinate system. When amplitude is also considered to be significant, or the primitive basis is normalized in magnitude, one tone may provide the referent for both phase and magnitude, and two further tones are sufficient to form a spanning set and have one component excess to a basis.

When two separated channels are considered, one tone in one channel may serve as referent, and the same tone in the second channel provide two referred elements, forming a basis for a two-space, or plane. To provide a spanning for a three-space, one further tone is necessary and also have one component excess to a basis. In neither case is a single tone capable of spanning (single channel or two channels). If three channels, suitably disposed, non-collinear, were available, it is conceivable that a single tone could be localized; however, there would remain an ambiguity as to which side of the plane the source was located.

Appendix C. Considerations of phase

It has been reported by Cramer & Huggins (1948) that sensation of pitch is produced when white noise is played by a headphone into one ear, and noise from the same source, modified only in phase relationships played through another headphone into the other ear. This was also verified by the author, and a further experiment undertaken to determine if phase conditions could be detected monaurally.

In order to produce changes in phase only, a non-planar network of the transfer function given in equation (C. 1) is employed

$$\frac{e_o(\omega)}{e_i(\omega)} = \frac{R(\omega) - jI(\omega)}{R(\omega) + jI(\omega)}. \tag{C. 1}$$

$e_o(\omega) \stackrel{\Delta}{=}$ output function of frequency,

$e_i(\omega) \stackrel{\Delta}{=}$ input function of frequency,

$R(\omega) \stackrel{\Delta}{=}$ the real part of the transformation numerator (same in denominator),

$I(\omega) \stackrel{\Delta}{=}$ the imaginary part of the transformation numerator (same in denominator),

$j \stackrel{\Delta}{=} +\sqrt{-1}$.

In order to realize such a network, figure 13 shows the circuit

$C \stackrel{\Delta}{=}$ a capacitor,

$L \stackrel{\Delta}{=}$ an inductor,

$R_L \stackrel{\Delta}{=}$ the intrinsic resistance of the inductor,

$R_e \stackrel{\Delta}{=}$ an external, or load, resistor.

The following equation applies

$$e_0 = e_1\left[\frac{R_e}{(R_e+R_L)+j\{\omega L-(1/\omega c)\}}\right]-e_2$$

$$= \frac{e_1 R_e - e_2\,[(R_e+R_L)+j\{\omega L-(1/\omega c)\}]}{(R_e+R_L)+j\{\omega L-(1/\omega c)\}}. \tag{C. 2}$$

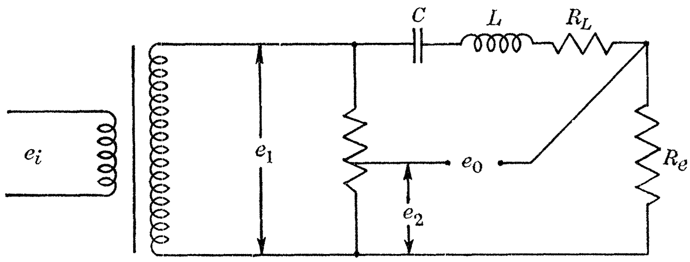

FIGURE 13. Phase shift only circuit.

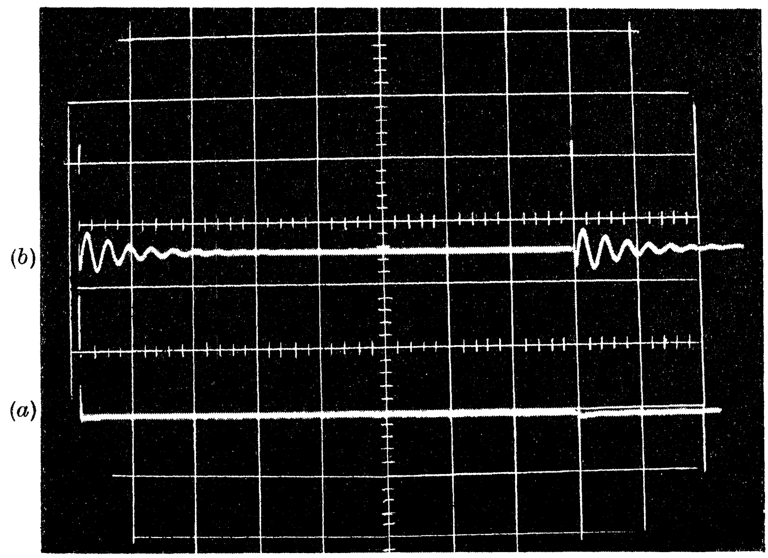

FIGURE 14. The input pulse (*a*) and the output (*b*) through the phase shift only circuit.

We adjust R_e or e_2, so that the condition of (C. 3) is met

$$2e_2(R_e+R_L) = e_1 R_e. \tag{C. 3}$$

Upon substituting (C. 3) into (C. 2), we obtain

$$e_o = \frac{e_2[(R_e+R_L)-j\,\{\omega L-(1/\omega c)\}]}{(R_e+R_L+j)\{\omega L-(1/\omega c)\}}. \tag{C. 4}$$

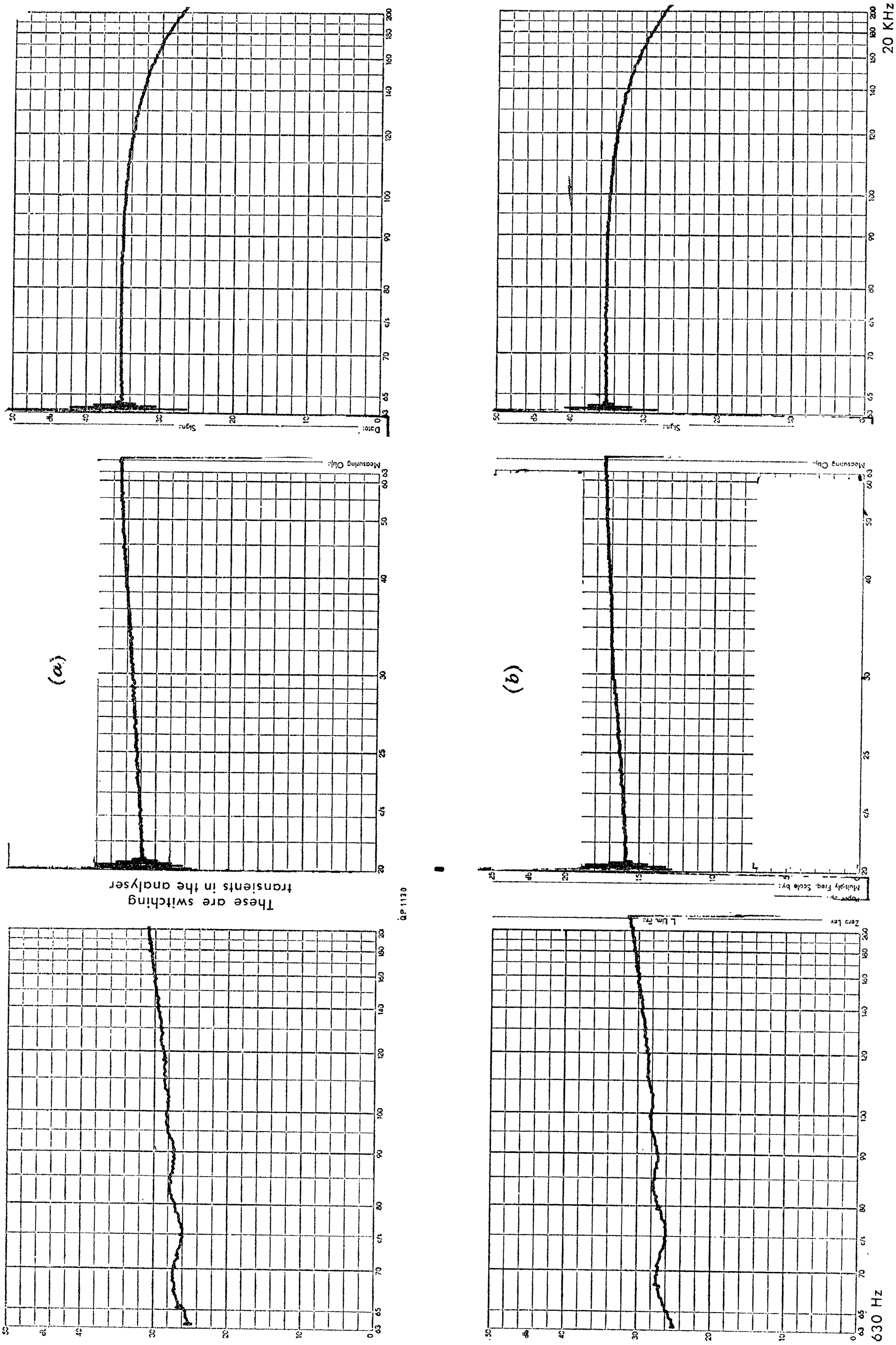

Figure 15. Spectrogram of pulse. (*a*) Pulse only; (*b*) Pulse through PSO, 1000 mm/s writing speed, 630 Hz to 20 kHz.

Within the frequency range of interest, the transformer enters merely as a ratio

$$\left.\begin{aligned} e_2 &= Ke_i, \\ K &= \text{a real number.} \end{aligned}\right\} \tag{C. 5}$$

Thus,

$$\frac{e_0(\omega)}{e_1(\omega)} = K\frac{(R_e+R_L)-j\{\omega L-(1/\omega c)\}}{(R_e+R_L)+j\{\omega L-(1/\omega c)\}} \tag{C. 6}$$

$$= K\frac{R(\omega)-jI(\omega)}{R(\omega)+jI(\omega)}.$$

Equation (C. 6) has a constant modulus (K) independent of frequency, but a phase shift dependent on frequency given by

$$\arg(\omega) = 2\arctan - \frac{I(\omega)}{R(\omega)}, \tag{C. 7}$$

$$0 \leqslant \arg(\omega) \leqslant 2\pi.$$

Figure 14 shows the comparative result of applying a pulse to such a circuit. (Figure 14*a* is the pulse as e_i and (14*b*) is the output e_0. Figure 15 shows spectrograms taken across the region of the maximum phase shift, 3 kHz, and only a tiny difference is observable by comparison (the result of the difference between the exact mathematical expression and the adjustment of the circuit). The difference can be heard, not as striking as the oscillogram, but capable of some interesting effects both monaurally and binaurally.

References

Angell, J. R. & Fite, W. 1901 The monaural localization of sound. *Psych. Rev.* **8**, 225–246; 449–458.

Batteau, D. W. 1961 Characteristics of human localization of sound. Final Report, II, United Research, Inc. under U.S. Navy Contract N123-(60530)23545A.

Batteau, D. W. & Plante R. L. 1962 The mechanism of human localization of sounds with application in remote environments. Final Report, United Research, Inc. under U.S. Navy Contract N123-(60530)27872A.

Bell, A. G. 1880 Experiments relating to binaural audition. *Am. J. Otology* **2**, 169.

Courant, R. & Hilbert, D. 1953 *Mathematical physics.* New York: Interscience Publishers.

Cramer, E. M. & Huggins, W. H. 1948 *J. Acoust. Soc.* **30**, No. 5.

Freedman, S. J., Stampfer, K. 1964 The effects of displaced ears on auditory localization. Institute for Psychological Research, Tufts Univ. under U.S.A.F. Contract AF-49(638)-1282.

Hartley, R. V. L. 1919 *Phys. Rev.* **13**, 373–385.

Hocart, A. M. & McDougall, W. 1908 *Brit. J. Psychol.* **2**, 386–405.

Kock, W. E. 1950 *J. Acoust. Soc.* **22**, No. 6.

Langmuir, I., Schaefer, V. J., Ferguson, C. V. & Hennelly, E. F. 1944 *A Study of the binaural perception of the direction of a sound source.* New York: G. E. Research Laboratory, Schenectady.

*Mach, E. 1874 *Archiv fur Ohrenheilkunde.* **7**, 214–226.

McLean. 1959 'Ears in action', a motion picture by U.S. Naval Ordnance Test Station, China Lake, California.

McNay, T. M., Stiles, L. E., Yaffa, E., Ware, V. E., George, A. & Baker, R. 1961 *Intelligibility experiments in reverberation.* Tufts Univ. Mech. Engng Dept.

*Rayleigh, Lord 1876 *Trans. Mus. Ass.*

* Quoted by Thompson (1881)

Rayleigh, Lord 1907 *Phil. Mag.* (6), **13**, 214.
Steinhauser, A. 1879 Phenomena of binaural audition. *Phil. Mag.* (5), **7**, 181, 261.
Thompson, S. P. 1879 *Phil. Mag.* (5), **8**, 385.
Thompson, S. P. 1881 On the function of two ears in the perception of space. *Rep. Brit. Ass.* pp. 716–7.
Wallach, H., Newman, E. B. & Rosenziveig, R. 1949 *Am. J. Psychol.* **62**, no. 3.
Wiener, N. 1949 *Time series*, New York: Wiley.
Wilson, H. A. & Myers, C. 1908 *Brit. J. Psychol.* **2**, 363–385.