NOTS TP 3109
Part 3
COPY 5

LOCALIZATION OF SOUND

Part 3. A New Theory of Human Audition

by

United Research, Inc.
Cambridge, Massachusetts

ABSTRACT. Localization of sound by humans has been shown to depend on a transformation of incident sounds by the pinnae, or external ear. The ears function as a computer-steerable array similar to an electronically swept radar antenna. The form of transformation is that of time delays. Autocorrelation of the time delays by mental function provides localization.

It has been found that the ability to localize sounds in another environment may be reproduced using microphones adapted with ear replicas and high-quality condenser headphones. Extension of this technique to underwater use has been effectively demonstrated, despite some component shortcomings. The experimental highlights that support the theory and the field tests to evaluate the operational utility of localizing systems are discussed.

The basic concept of autocorrelation of time delays introduced by the pinna has been extended to speech recognition problems. A new theory of human audition, which ascribes significance to the time domain rather than the frequency domain, has been developed that explains not only binaural and monaural localization, but also the "cocktail party effect," pitch discrimination, speech recognition, masking, intelligibility in reverberation, and other auditory phenomena.

U.S. NAVAL ORDNANCE TEST STATION
China Lake, California
December 1963

54

U. S. NAVAL ORDNANCE TEST STATION
AN ACTIVITY OF THE BUREAU OF NAVAL WEAPONS

C. BLENMAN, JR., CAPT., USN — *Commander*
WM. B. MCLEAN, PH.D. — *Technical Director*

FOREWORD

The purpose of this report is to describe results of studies in sound localization carried out over a 3-year period. A new theory of human audition has been developed from localization studies and is presented here.

This work was conducted by United Research, Inc., Cambridge, Mass., under Contract No. N123(60530)30383A issued by the U.S. Naval Ordnance Test Station. The information herein covers work that completed this contract in September 1963.

This is Part 3 of a continuing series of reports that will be issued covering various aspects of the subject. The text pages in this report were reproduced in facsimile from a report issued by the contractor.

WM. B. McLEAN
Technical Director

NOTS Technical Publication 3109, Part 3

Published by	Publishing Division Technical Information Department
Collation	Cover, 15 leaves, abstract cards
First printing	130 numbered copies
Security classification	UNCLASSIFIED

CONTENTS

Chapter 1 Introduction 1
2 Theory of Sound Localization 5
3 Experimental Results 11
4 Theory of Human Audition 17
References 28

CHAPTER 1

INTRODUCTION

1.1 Introductory Remarks

This report is a summary of the studies conducted by United Research Inc. for the U. S. Naval Ordnance Test Station on the subject of sound localization. To date significant discoveries have been made from which a new theory of human audition has evolved. It is expected that this work will have far reaching effects in both scientific and commercial areas.

A person's ability to localize sounds in his environment has occupied the attention of many researchers for a century. Despite this span of time and the efforts expended therein the factors explaining this taken-for-granted phenomenon resisted discovery. However, as the science of information theory progressed, new insight into physical behavior was gained which today adequately defines the functions associated with hearing, namely, localization, attention, and recognition. Prior to this the basic result, despite the voluminous data published, was that localization of sounds is accomplished by a mental discrimination of intensity and/or phase differences between the two ears. The enormity of information supporting this result had all but created a complete deterrent to further investigations of localization based on different concepts. Nevertheless, other questions relating to localization still remained unanswered. For example, how is the median plane ambiguity resolved? How is monaural localization accomplished? How does a person pay attention to a particular sound among many, the cocktail party effect? How is pitch discrimination achieved? These and other questions are not answered within the framework of current theories of hearing. Spurred on by a demonstration of localization to be described, research was undertaken which has led to discoveries in human hearing which have shown consistency and have provided explanations for the observed phenomena related to hearing.

The first section of this report will describe the developments supporting the original hypothesis concerning localization while the second section presents a new theory of human audition. The planned extensive experimental program to test the theory has not yet been carried out. However, each facet of the theory has been supported in brief experiments. This is in no way an apology for lack of supporting evidence. Rather, it is indicative of the extensive work and time which has not yet been found available. It is felt that the uniqueness of the theory will excite other researchers with the time and facilities to undertake further validation. When the full experimental program has been completed it will be the subject of a scientific paper for publication.

1.2 Historical Background

At a summer seminar held at Harvard University in 1959, Dr. William B. McLean, Technical Director of the U.S. Naval Ordnance Test Station, China Lake, California, demonstrated to Dr. D. W. Batteau of United Research Inc. that distorting the pinnae, or external ear, changed the subjective perception of the location of a source of sound. With his ears undistorted and his eyes closed, Dr. Batteau was asked to point to the location of keys jingled by Dr. McLean. Localization was accurate. With his eyes closed, Dr. Batteau then distorted his ears by pushing the concha of his ear from behind with the first finger of each hand. The ear canal opening was not disturbed. Responses to the jingling keys were now almost always incorrect, partiqularly in elevation. Further work in this area was done by Mr. Robert Cohen, a graduate student at Drew University, which confirmed the role of the ear in sound localization. With the support of NOTS, United Research undertook a program to study the demonstrated phenomenon.

1.3 The Hypothesis – A Statement of Expectations

The cause and effect relationships described prompted a view of the situation from the standpoint of information theory. In aural perception the interplay of sound source, medium of transmission, receptor characteristics, and interpretation or recognition have all at some time been assigned sufficient importance to explain observations. However, it can be shown that the sound source and the medium do not provide localization information since such sound reproduced from a microphone is not localizable. The other factors do play a role as suggested by the demonstration.

In accordance with information theory knowledge is derived by inverting a transformation introduced on the observed space by the mechanism of perception. Hence, to know the location of a sound requires a transformation of the acoustic space by the auditory mechanism and the inversion of this transformation by mental function. A hypothesis was stated that the transformation of the acoustic space was introduced by the pinna, or external ear. The form of this transformation was that of a series of delays of the incident sound enroute to the eardrum caused by diffractions and reflections in the ear. Because of the asymmetrical configuration of the ear, this transformation is unique for every aspect of ear and sound source in the half-space of the ear. The inverse of this transformation is formed by autocorrelation of these delays by the mental function. By this same inverse attention is payed to the point localized. It will be shown that this hypothesis, subjected to experiment, provided substantiating results, and furthermore provided answers to the questions noted above consistent with the functioning of the hearing mechanism reported by Von Bekesy and others.

1.4 Popular Theory of Human Localization

By way of background, the currently accepted theory of localization will be briefly reviewed. Since man has two ears, it is obvious that orientation of these ears in the wave front will give some directional information. It is conceivable that a mental process may be used to measure the time, or intensity, differences of a sound at each ear. The idea of "phase" difference has provided by far the most popular hypothesis for localization. By logical necessity, a receptor system providing the information of time differences determines a plane in which the sound source lies. To provide a three coordinate space the head, or ears, therefore must be moved to two linearly independent positions, to determine two additional planes. The intersection of these planes locates the sound source. Since translational motions of the head are not frequently observed, rotation of the head about the ear-ear axis, the neck-head axis, and the nose-rear of the head axis is required to produce three linearly independent positions. For reasons to be explained, a steady tone is localized in this manner with associated gross head motions. But transient sounds are localized without head motion and furthermore can be localized monaurally The underlying logic applicable to these observations is now understood.

CHAPTER 2

THEORY OF SOUND LOCALIZATION

2.1 Logical Structure of the Theory

The initial logical statements concern dimensionality. To define a point in three-space requires a three-dimensional coordinate system. At least four points are necessary to determine a three-space: one point at the origin and one point on each coordinate line. If the signal is two dimensional, i.e., a pure tone, the head is observed to move to provide the required dimensional information.

When one ear localization is considered, the spanning requirements apply to the signal: the signal must be sufficiently complex so as to provide a coordinate system. This requires at least four linearly independent components in the signal. In addition there must be a transformation which transforms each of the components differently onto the sensory point.

When the spanning and transformation requirements are met, the next logical requirement is the existence of a unique inverse of the transformation for each point which can be localized. This inverse must reconstruct the signal at the origin from its transformed character; the same inverse "pays attention" to the point localized.

Finally, we relax our transformation and inverse requirements to provide a finite resolution. The inverse may then be approximate and not necessarily exact, but must maximize a measure for a particular point localized or to which attention is paid. The maximization must then be in terms of the information rate or channel capacity to that point. This may also be stated as minimal temporal redundance for that point.

2.2 Mathematics Applicable to the Theory

The simplest mathematics applicable to sound localization, so far as the transformations and coordinate systems are concerned, are those of Hilbert Space at its simplest. The mathematics applicable to processing of the received signal, localizing, and paying attention are those, in addition, of Information Theory.

Logically, the signal may be expressed by equation (2.2-1)

$$S(\rho, \theta, \beta, t) = \sum_{i=1}^{n} s_i (\rho, \theta, \beta, t) \tag{2.2-1}$$

$$S(\rho, \theta, \beta, t) \triangleq \text{the sound at the source} \tag{2.2-2}$$

$$\rho \triangleq \text{distance from observer to sound source} \tag{2.2-3}$$

$$\theta \triangleq \text{azimuth angle from observer referrent to sound source} \tag{2.2-4}$$

$$\beta \triangleq \text{altitude angle from observer referrent to sound source} \tag{2.2-5}$$

$$t \triangleq \text{time} \tag{2.2-6}$$

$$s_i \triangleq \text{components of sound} \tag{2.2-7}$$

$$n \geq 4 \tag{2.2-8}$$

There is also a transformation, T, which varies with the orientation of the observer and the sound source

$$T(\rho, \theta, \beta) \triangleq \text{transformation of incident sound by observer} \tag{2.2-9}$$

The perceived signal, P, is expressed by equation (2.2-10)

$$P(\rho, \theta, \beta, t) = T(\rho, \theta, \beta)\, S(\rho, \theta, \beta, t) \tag{2.2-10}$$

If we now assume that the information density in a signal cannot be increased by a transformation, T, then the transformation, M, which produces the maximum information rate is the one to apply to the perceived signal in order to determine the localizing transformation. Since

$$C = f \log_2 (1 + \frac{p}{n}) \qquad (2.2\text{-}11)$$

$$C \triangleq \text{channel capacity} \qquad (2.2\text{-}12)$$

$$f \triangleq \text{bandwidth} \qquad (2.2\text{-}13)$$

$$p \triangleq \text{information power} \qquad (2.2\text{-}14)$$

$$n \triangleq \text{noise power} \qquad (2.2\text{-}15)$$

then

$$[M]\ [P] \longrightarrow C_{max} \qquad (2.2\text{-}16)$$

determines the transformation, M, corresponding to the localizing transformation, T. In general, the inverse, T^{-1}, to a transformation T produces infinite bandwidth, and hence maximum channel capacity. However, T^{-1} is in general impossible, so that we may write equation (2.2-17)

$$M \doteq T^{-1} \qquad (2.2\text{-}17)$$

Thus the mental operation of maximizing channel capacity is equivalent to finding the particular inverse to the transformation providing localization.

The second factor in the channel capacity equation may also be considered, by comparing localizing maxima against unlocalized values.

$$\frac{\sum_{i=1}^{n} \left\{[M]\ [T(s_i)]\right\}^2}{\sum_{i=1}^{n} \left\{[1]\ [T(s_i)]\right\}^2} \doteq \text{maximum} \qquad (2.2\text{-}18)$$

The "peakedness" of the signal is under consideration here, for maximum peak factor may give maximum power. Consider the sum of two separated unit voltage pulses against two unit voltage pulses superimposed which gives a "power" ratio of 1:2.

From the information theory point of view both bandwidth and peak factor are significant in localization. A reasonable hypothesis that power peaks and bandwidth are maximum in the primitive signal points to the inverse of the localizing transformation. If the inverse is known, the point of origin is known, so that this is a means of inferring the inverse. Since the inferring assumptions need not be universally true, the ear can be fooled, and experimental hallucination provides a good test of the hypothesis.

2.3 Transforming Mechanisms

The mechanical equipment available for transformations are (1) apertures and (2) paths. These may also be stated as selective delay mechanisms. Previous experience with directional couplers suggested an analogy between the function of the pinna and that of a directional coupler.

A one-dimensional directional coupler, as used with a wave guide (early WW II development) is sketched in Figure 2.3-1. A wave propagated in the direction B has coupling through the holes in phase at the probe. Thus only the waves of direction B are sensed.

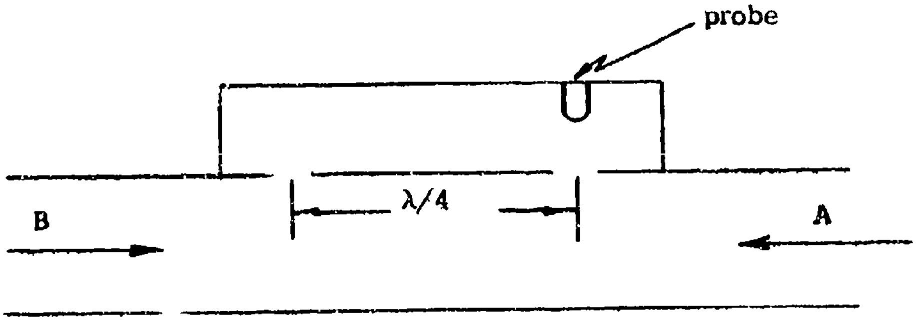

Figure 2.3-1 One-Dimensional Directional Coupler

If the coupler is considered as a two-dimensional device with incident pulsed wave fronts, as sketched in Figure 2.3-2, then the time spacing between the two pulses tells the direction in the half plane of the incident wave front, Autocorrelation of the three signals would show maxima at the different relative delays for the three pulses at differing angles of incidence. Peak factors would be also maximum for differing delay additions corresponding to maxima in the autocorrelation functions.

If a coupler sensitive to both azimuth and elevation is desired, three holes may be put in a plane, with an internally delayed path for one pulse domain to keep it ever separated from the other domain. This configuration is sketched in Figure 2.3-3. After making measurements on an ear, such a three-hole coupler was built and tested with gratifying results. Direction location in space was possible with the couplers. Although the mechanism of the pinna is more complicated than that of the coupler, the experimental evidence for its role in sound localization is significant.

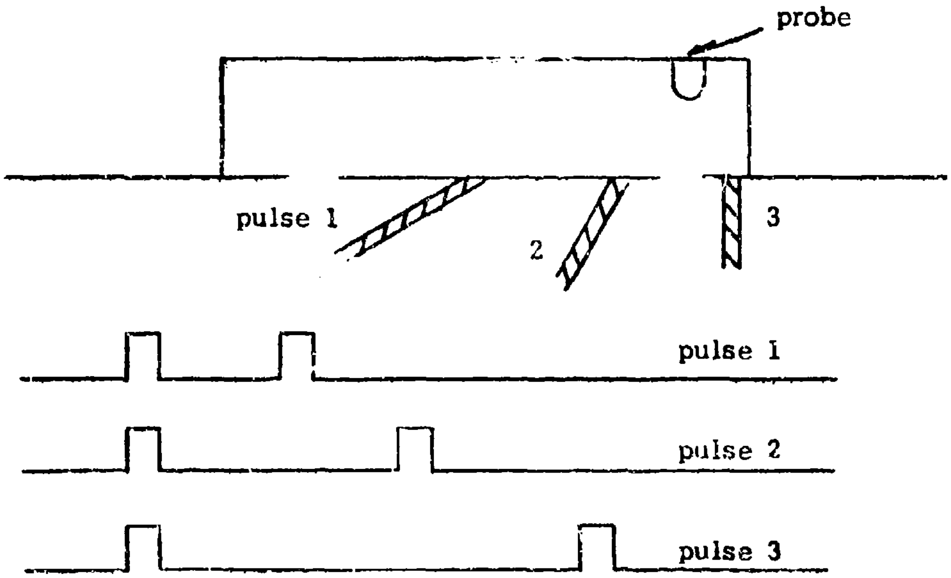

Figure 2.3-2 A Directional Coupler in a Plane

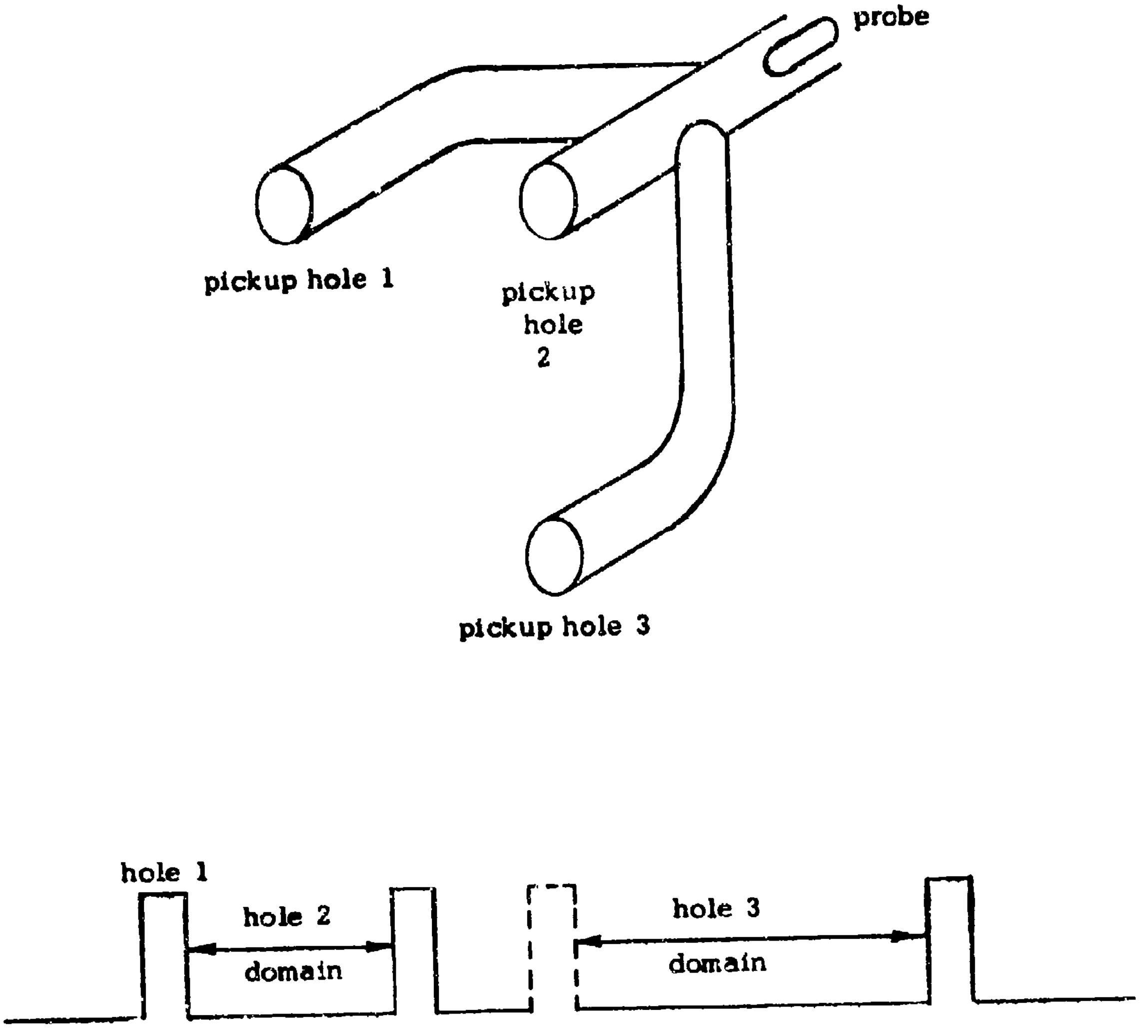

Figure 2.3-3 A Three-Hole Coupler for Azimuth and Elevation

CHAPTER 3

EXPERIMENTAL RESULTS

3.1 Early Experiments – Observation of Time Delays

The progress of research in localization may be tied directly to a few basic experimental discoveries.

a) Confirmation of time delay behavior for different aspects of ear and sound source.

b) Importance of high quality microphones, headphones, and electronics in aural coupling.

c) Effects of acoustic properties of materials.

d) Acoustic isolation of the microphones from all sound transmission paths except the ear canal.

e) Other supporting observations.

Time delay behavior was confirmed using a plaster ear replica five times normal size and an AKG C-26 condenser microphone. A pulsed electrostatic speaker was used for the sound source. The ear was mounted at the center of an aluminum sheathed board, 8' x 4'. The speaker was moved in an azimuth and elevation semicircle about the ear. Figure 3.1-1 is a tracing of the oscilloscope photograph showing the change in delay with azimuth. Figure 3.1-2 is a graphical representation of changes in both azimuth and elevation.

An enlarged ear was used to improve resolution. Subsequent work has included measurements in a Freon atmosphere in which sonic velocity is about half that in air. To date, however, adequate resolution to cover the entire half-space has not been achieved

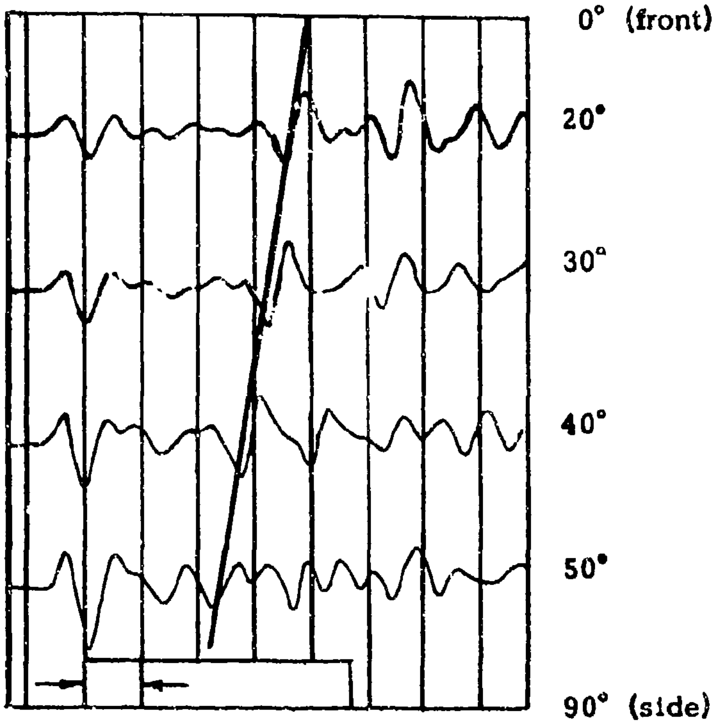

Figure 3.1-1 Oscillogram Tracing of Azimuth Delay Changes

3.2 Localization Reproduced

A system was developed to permit a subject to be aurally coupled to another environment whereby he might perceive it as though he were in it, The device consisted of a mannequin head fitted with rubber ear replicas behind which were placed Electrovoice 649A microphones. When coupled with a high fidelity amplifier and a good quality headset, externalization was subjectively evident. Localization was possible with difficulty. Prior to this no significant importance was attached to component quality. As a result electronic equipment of original design was built and AKG C-26 condenser microphones purchased. At the listening end Beyer DT-508 semi-insertion headphones were bought to eliminate a second transformation by the listener's ears. This array of equipment was a big improvement for now sounds could be localized above, below, behind, and with some ambiguity in front. Recordings were made with this system and demonstrated the superior quality of the recorded sound,

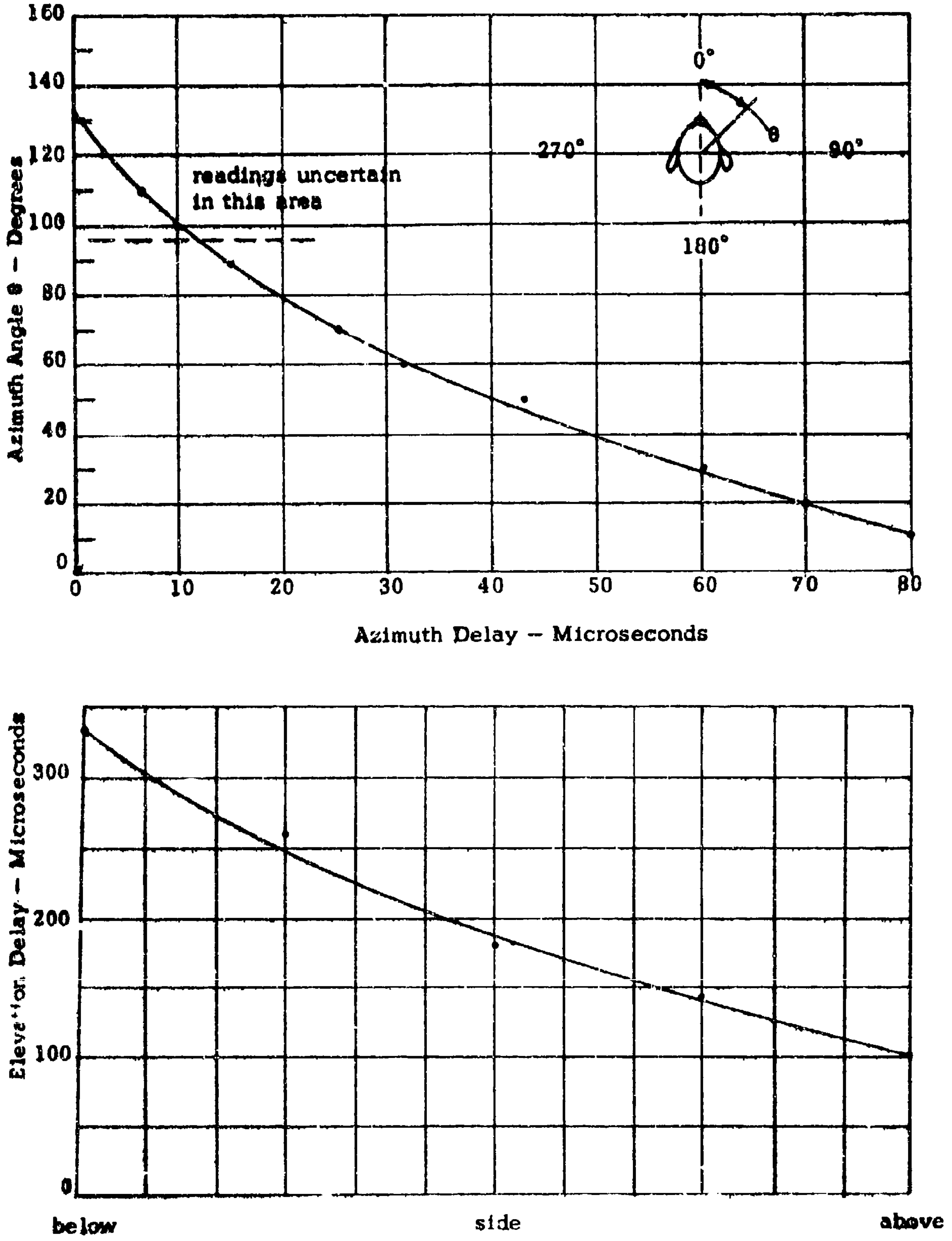

Figure 3.1-2 Delays vs. Orientation

Since the ultimate standard of performance is localization with one's own ears, improvements in microphones and headphones were still sought. Bruel & Kjaer condenser microphones (Mod 4133) were found which have response characteristics to 40,000 cycles from 40 cycles within ± 1.5 db. A comparable headphone was out of the question. Out of desperation, a headset was made from two condenser microphone cartridges (Mod 4135). The system consisting now of rubber ears, B&K microphones and condenser headset permits localization after a minimum familiarization time which nearly duplicates human ability.

A few remarks are required regarding the need for equipment fidelity beyond the audio range. Published data shows that the sound pressure at the eardrum compared to that before the ear is very small. Thus one suspects that components of inaudible intensity provide no infomation. From the standpoint of localization, this is true. However, broad bandwidth is necessary to resolve the delay times to permit autocorrelation by the mental function. It is this resolution capability which preserves redundance that permits the reproduction of localization and the recording of 'actual presence.'

3.3 Acoustic Properties of Materials

The human ear in air is an acoustically hard substance. Since we were anxious to improve our system in every respect, we tested a rubber ear and found that for frequencies above 8,000 cps it is acoustically soft. To avoid the problems of casting ears of metal, such as brass, a dense, high elastic modulus castable epoxy was used. Tests show an improvement in coloration in the higher frequencies which is expected to further improve the localization system.

3.4 Acoustic Isolation

It was recognized early that the microphones behind the ears must be isolated from all acoustic transmission paths except the ear canal. While some success was realized in this regard, the problem of complete isolation remains, Effort is underway to suspend the ear and microphone in a dielectric gel in an effort to reduce transmissibility and improve acoustic isolation.

3.5 Other Observations

During the conduct of research certain brief but interesting experiments were carried out yielding results to support the new concepts. One such experiment involved filling only a subject's ear canal with water. The thought was that each delay introduced by the ear would be increased in velocity enroute to the eardrum by the same amount with no net change in the relative delays. Monaural localization under this condition was accurate. When part of the ear was filled with water, it was expected that each delay would be changed with the net result that the relative delays would be different and localization distorted. This was found to be the case.

In a second experiment, a rubber ear was placed backwards over one ear. With the other ear plugged, localization was reversed. This effect was duplicated many times and provided a quick demonstration of the effect of the ear.

3.6 Applications

A major effort which occupied the period covered by Contract No. N123-(60530) 30283A involved the development of a system to permit localization underwater. The purpose was to test the short range navigational utility. An obvious benefit in this application is the use of the operator's own mental faculties for locating unknown sounds.

Sonic velocity in water is 4.5 times that in air. Hence, to preserve the same time delays used in air, the ears for use underwater were enlarged 4.5 times. For proper acoustic mismatch they were cast in stainless steel and an oyster-type hydrophone isolated by a steel block mounted behind each. The ears were fastened to the ends of 3" diameter tubing separated by 43 inches. A spherical covering of Absonic A was provided to minimize flow noise.

This system was suspended from a motorized barge 30 feet deep at Morris Dam. A mechanical noise maker was lowered to various depths at different distances from the barge. The subject listening through the pickup indicated the direction the operator should proceed. Each subject was covered so that no visual references were possible. With the sound source fixed, the barge was navigated to it every time over distances to 500 yards. In some tests, the sound source was taken to a cove outside of visual range. Again, it was found using only acoustic information.

These tests have further confirmed earlier results in underwater localization. Although we were limited to a two-dimensional coordinate space, it is felt that use of this device on vehicles having three-dimensional capability in water would be a distinct asset.

CHAPTER 4

THEORY OF HUMAN AUDITION

As a development of our research in localization, Dr. D. W. Batteau extended the basic concepts to other areas of human audition to provide a many faceted theory which displays elegant consistency. We set out to examine the means by which man localizes sound. We found that two ears are not necessary for localization, for man can do it with one. We also found that the pinna, or external ear, has an essential role in this perception. By this means man can localize and can also pay attention to a selected locale. When the process was examined with respect to the nervous system, it was possible to construct new models in which the known structure was well suited for producing the required function. In addition, from this study came an understanding of function and structure which permitted an examination of human speech recognition with rewarding results. The evolution of this study is presented in the following paragraphs without inclusion of the experimental details as noted early in this report.

4.1 Introduction

The hearing function can be separated into several, not necessarily independent, functions. These functions are: (1) localization, (2) attention, and (3) interpretation. The function of localization is to assign a place of origin to sound. The function of attention is to select from a mixture of sounds one or a group which has particular importance. The function of interpretation is to assign meaning to the sound or to the one to which attention is paid.

A mechanism is needed which will provide for the three functions since their existence may easily be demonstrated, which can be stated logically and to some extent mathematically. A logical point can be made concerning localization: since the point of origin is located in a three-dimensional space, a means of establishing a three-dimensional coordinate system is required. A second point can be made concerning attention: from an information theory viewpoint, the audible message to which attention is paid must be redundant in a unique way. A third point can be made regarding selection: an invariance in characteristic of the selected message must be established. To meet scientific requirements, all of these logical requirements of function should be met simply and in a manner suited to the evident organization of the auditory sensory system.

4.2 Historical Background

From a survey of literature regarding hearing we find that there have been well defined mainstreams of thought and experiment. Localization research has concentrated on the obvious fact that man has two ears (Ref. 1) although some work has been done in monaural localization (Ref. 2). Research concerning selection has been based principally on a mechanical model of the inner ear involving the cochlea and the basilar membrane as a resonant structure (Ref. 3).

The binaural approach to localization first fails to meet the logical requirements for a three-dimensional coordinate system if it is considered simply as two separated perceptors discriminating time or intensity. It can be shown that for a difference in time of arrival of signals at the two ears given by the following equations:

$f_1(t)$ = sound at ear no. 1

$f_2(t)$ = sound at ear no. 2

and

$$f_1(t) = f_2(t + T)$$

where T is the time difference

then a pair of hyperboloidal surfaces exist, with every point on one surface the possible origin of the sound. It is obvious that similar surfaces are defined for any given relative intensity between the two ears. These two functions can then at most determine lines of intersection of the two families of surfaces for any given value of time difference and loudness difference. A three-dimensional coordinated system in either or both quantities can be generated by moving the pickup pair into successive, non-collinear positions, and research has shown that head motions do have a role in localization. A coordinate system can be generated either by translation or rotation of the head, but of these only rotation appears to have any practicality. Turning the head will generate a coordinate plane and nodding from side to side will generate another. A third may be generated by nodding up and down. It is conceivable that a combination of time, intensity, and head motion could indeed generate a coordinate system which would permit the establishment of the point of origin of a sufficiently sustained sound. Further, in view of the logical structure described earlier, signals which do not provide a spanning set are localized in this manner.

The evident defects in the experimental value of the currently popular theories are three: (1) no mechanism is provided for the mental function of attention; (2) no mechanism is provided for localization without head motion (transient sounds); (3) no mechanism is provided for monaural localization. These three conditions have all been well demonstrated, however.

Notable work has been done in physiology relative to the cochlea, the basilar membrane, and the associated system of nerves (Ref. 4). It has been theorized that the change of shape of the basilar membrane provides a frequency selection mechanism which selects tone. The observed sharp discrimination of tones compared to the observed broad flexing of the basilar membrane has occupied the attention of some researchers. In this theory and its adjuncts there are several defects: (1) the sharp discrimination of tone is not well explained; (2) the sensing of coloration of noise is even less satisfactory; (3) no mechanism of attention to low or high tones, or to voices by some mental function is provided; (4) no mechanism of selection or identification is provided.

4.3 Demonstration

It is easy to demonstrate that the external ears have a role in localization by simply distorting them. If a subject with good hearing pushes in the concha from behind with his fingers, it will severely alter his perception of altitude. Persons capable of accurate altitude sensing both high and low will be influenced towards the middle position by this deformation. More detailed experiments have been conducted to show that distorting or negating the pinna produces deviations in normal accuracy of localization. Perhaps the best illustration is to reverse one's ear, front to back by means of a rubber replica and witness the reversal in monaural localization.

4.4 Initial Theory

As a result of these demonstrations a hypothesis was formed that the external ear performs a mechanical transformation on the incoming sound waves which could be used by mental processes for localization. In a literature search we found one reference to monaural localization, Ref. 2. It suggested to us that the external ear performed a spanning transformation on the incident sound. We need not then consider both ears as essential to the process, but rather consider them as reinforcing elements in the system. In the remainder of the literature the pinnae were taken as functionless or as sound gathering devices.

After a series of experiments we concluded that the external ear introduces several delay paths in the route of the sound wave to the ear drum. The mental process should then be that of autocorrelating the resultant signals to locate the delays and thus ascertain the point of origin.

A consequence of signal processing of this sort should be not only localization but also enhancement of the signal localized by the correlating process. The process of correlation makes use of the redundance in the signal produced by the several paths. This redundance, used in correlation, provides an improvement in signal to noise ratio of the particular signal and thus provides a mechanism of attention as well as one of localization.

If time delays were significant, the correlation should be invariant under most real time operations, in particular that of differentiation. A pickup system consisting of a mannequin head adapted with ear replicas behind which were mounted high quality microphones was used in conjunction with broadband electronics and the best commercial headphones available showed that once and twice differentiated signals were localized as well as the original. This particular theoretical aspect provided a suggestion for extension of the theory.

4.5 Extension of the Theory

Since it has been shown that the localization and attention mechanisms were provided by time correlation of a transformed sound, it seemed possible that speech provided a similar situation. In this case, the transformation of the vocal pulse or other spanning sounds (regurgitated air or an artificial larynx) would be provided by the vocal tract. The recognition process would be that of locating the delay times, or redundances, to infer the shape of the vocal cavity.

The recognition of speech should then remain invariant under most real time operations. By extending the work done by Licklider (Ref. 5), it was shown that once and twice differentiated speech was indeed as intelligible as the original. Furthermore, the correlation points of the second derivative corresponding to the maximum information rate in the original signal should lie on the zero axis and clipping should alter intelligibility very little, if at all.

Tests with twice differentiated, clipped speech showed that intelligibility was not impaired. It was further reasoned that if the time correlations were of principal importance, then filtering the twice differentiated clipped speech should mark these points as the initiation of the characteristic pulse response of the filter. Drastic filtering of the signal was shown to retain a large measure of intelligibility.

4.6 Further Theory

While the results predicted by theory had been achieved experimentally, the mechanism of correlation remained unexplained. A hypothesis was stated that the nerves on the basilar membrane and the subsequent neural networks act as delay lines and that correlation could be formed by tapping along them with computational nets. If this were the case, a short delay could be correlated at the beginning of the basilar membrane, but a long delay would require up to the total length of the membrane.

While we were not able to experiment in this area directly, reference to the work of Von Bekesy (Ref. 4) showed consistency. His work showed that high frequencies (short correlations) were perceived at the initial part of the basilar membrane, while low frequencies (long correlations) were perceived at the final part.

4.7 Continuation of Theory

All of the hypotheses had been supported by experiment with a consistency that was gratifying, but the computational scheme was still missing. We then formed the hypothesis that the auditory system would function on a basis similar to that indicated by the information theory equation.

$C = B \log_2 (1 + \frac{s}{q})$

C = bits per second

B = bandwidth in cycles per second

s = signal power in watts

q = quantification power in watts

In this case the measure C, being one of information intensity, could be equated with loudness.

If the hypothesis were true, subjective loudness should increase directly with bandwidth for equal total power, and logarithmically with power for a fixed bandwidth. A search of the literature indicated that such might be expected (Ref. 6) and our own experiences provided confidence. Further tests are planned.

If the hypothesis were true, then the comparison of binaural loudness with monaural loudness should provide evidence, since the binaural bandwidth for identical ears is twice that of the monaural, without tonal change. Experiments showed that indeed a subjective loudness binaurally followed the expected pattern.

4.8 Integration of Theory

The hypothesis can be integrated in view of the consistent experimental results to provide a theory of human audition.

1. The sensory system responds to provide nerve signals following the rule

$$I = B \log (1 + \frac{s}{q})$$

2. Recognition by audition occurs by forming autocorrelations on the incoming signals for a single channel and cross correlations for two channels.

3. The basilar membrane serves as the initial delay element in forming correlations.

4. Subsequent correlations are formed in the continued nerve network, including the correlation of memory.

4.9 Prediction and Synthesis

From the theoretical considerations of human audition, it is possible to make experiments for which the outcome is predictable, and also to investigate previously reported experiments for analysis. One of the predictable phenomena is masking of one sound by another (Ref. 7).

4.10 Masking

Consider two signals having the same bandwidth but orthogonal correlation

$$L = B \log_2 \left(1 + \frac{s_1 + s_2}{q}\right)$$

$$\frac{dL}{d s_1} = \frac{B q}{q + s_1 + s_2}$$

This equation illustrates the dependence of comparative loudness of s_1 on the value of s_2. Consider two signals in orthogonal bands

$$L = B_1 \log_2 \left(1 + \frac{s_1}{q}\right) + B_2 \log_2 \left(1 + \frac{s_2}{q}\right)$$

$$\frac{dL}{d s_1} = \frac{B_1 q}{n + s_1}$$

This equation shows the relative lack of masking by the independence of s_1 and s_2.

4.11 Motion Synthesis

If the role of the pinna is to provide particular correlations for localization and attention, then a pair of pulse separated by lengths comparable to the delays produced by the pinnae should appear to move as the spacing is changed. In our simple experiments such was observed to be the case for variations in separation between 0 and 250 microseconds for short electronic pulse pairs perceived monaurally.

4.12 Intelligibility in Additivity

Single side band transmission of voice can result in the addition of a constant to all frequency components by inaccurate reinsertion of the carrier. If the correlation of time information provides intelligibility, then clearly marking the significant epochs should retain intelligibility even when the components are additively modified. An experiment using band limited twice differentiated, clipped speech shows that such is indeed the case. Even with frequency shifts well beyond those tolerable for formant frequency theory, high intelligibility is retained.

It may be concluded that the formant frequency theory of speech recognition is incorrect. It approximates the situation only to the extent that spacing of correlated phenomena imply frequency spectral density distribution components; the converse is not equivalent.

4.13 Reverberation

With several correlation domains identified, aural and vocal, it is reasonable to predict others, including what is ordinarily termed reverberation and also including memory. Signals reflected from walls could be used to improve the perception of sound from a particular location since the related transformation is unique for a particular relationship of sound source and hearer.

An experiment was designed to find thresholds outdoors compared to thresholds indoors for recognition of words from a PB word list. The expected results were obtained showing subjectively louder sounds indoors than outdoors.

4.14 Intelligibility in Redundance

The "coctail party effect" provided by the pinna transformation permits attention to be focused on sounds having a unique correlation function. This can be synthesized by electronic delay lines so that speech thus processed should be more intelligible than speech not thus processed in the presence of uncorrelated, random noise. The constraints of the experiment provided for equal signal power in the redundant and non-redundant conditions before addition. The results were somewhat as expected with a variability which may indicate modular construction of the nervous system.

4.15 Subjective Dynamic Range

Because of the differing slopes of binaural and monaural loudness curves given theoretically by the difference in the equivalent of bandwidth for two channels compared to one channel (a theoretical limit of twice the bandwidth), changes in signal power perceived binaurally are interpreted as greater changes in loudness than the same changes in signal power perceived monaurally. This effect is noticeable when listening to symphonic music of wide dynamic range with high fidelity headphones.

4.16 Subjective Bandwidth

The same arguments of binaural versus monaural subjectivity applied to dynamic range apply to bandwidth. Since small changes are more easily perceived binaurally, the perception of derivative scale is improved binaurally, where the time required to perceive a given slope is shortened. Two channel systems of equivalent bandwidth but separable correlations would then give the subjective impression of extension of the upper frequency range. Listening to sounds of all kinds, speech, ordinary noises and music provide the predicted results.

4.17 Auditorium Acoustics

From the theory and evidence it appears that the amount of time delay introduced by multiple paths is an important factor in indoor listening. An analogy between consonance and dissonance may be drawn for similarity in subjectivity between phenomena of one physical scale (the length of musical instruments) and another (the spacing of auditorium walls). It is theoretically insufficient merely to control the iterative factors, or rate of sound decay. Experimentation in this field has been beyond our means, but the evidence of listeners to symphonic music, plays, and other sonic performances is abundant.

4.18 Correction of Impaired Hearing – Otoscelerosis

Since the correlating intervals are invariant with differentiation, it seemed possible that impaired hearing of at least one type could be helped. Otosclerosis results in stiffening of the mechanical coupling between the ear drum and the oval window and should affect long-scale motions more than short scale. A brief experiment in this area showed that the existence of an intact basilar membrane could be inferred by motion synthesis and that perception and recognition could be brought to normal at normal power levels by differentiation (sharpening of the epochs) of the signal in the two cases studied, one of 60 db loss in one ear and one of 40 db loss in both ears. In both cases, introducing the pinnal redundances resulted in significant improvement in subjective coupling, or pleasure and reality in audition.

REFERENCES

1. Trimble, O.C., "Intensity Difference and Phase Difference as Conditions of Stimulation in Binaural Sound Localization," Amer. J. Psych. 47 [1935] .

2. Angell, J.R. and Fite, W., "The Monaural Localization of Sound," Psychol. Review 8 [1901] .

3. Neff and Yela, "Function of the Auditory Cortex: The Localization of Sound in Space," Amer. Psych. 3 [1948] .

4. von Bekesy, G., Experiments in Hearing, McGraw-Hill, 1960.

5. Licklider, J.C.R., "Effects of Differentiation, Integration, and Infinite Peak Clipping upon the Intelligibility of Speech," Journal of the Acoustical Society of America, Vol. 20, No. 1, January 1948.

6. Shaw, W.A., Newman, E. B., Hirsh, I. J., "The Difference between Monaural and Binaural Thresholds," Journal of Exp. Psych. 37 [1947] .

7. Kock, W.E., "Binaural Localization and Masking," J. Acous. Soc. Amer. 22 [1950] .

ABSTRACT CARD

U.S. Naval Ordnance Test Station
Localization of Sound, Part 3. A New Theory of Human Audition, by United Research, Inc. China Lake, Calif., NOTS, December 1963. 28 pp. (NOTS TP 3109, Part 3), UNCLASSIFIED.

ABSTRACT. Localization of sound by humans has been shown to depend on a transformation of incident sounds by the pinnae, or external ear. The ears function as a computer-steerable array similar to an electronically swept radar antenna. The form of transfor-

(Over)
2 cards, 4 copies

U.S. Naval Ordnance Test Station
Localization of Sound, Part 3. A New Theory of Human Audition, by United Research, Inc. China Lake, Calif., NOTS, December 1963. 28 pp. (NOTS TP 3109, Part 3), UNCLASSIFIED.

ABSTRACT. Localization of sound by humans has been shown to depend on a transformation of incident sounds by the pinnae, or external ear. The ears function as a computer-steerable array similar to an electronically swept radar antenna. The form of transfor-

(Over)
2 cards, 4 copies

U.S. Naval Ordnance Test Station
Localization of Sound, Part 3. A New Theory of Human Audition, by United Research, Inc. China Lake, Calif., NOTS, December 1963. 28 pp. (NOTS TP 3109, Part 3), UNCLASSIFIED.

ABSTRACT. Localization of sound by humans has been shown to depend on a transformation of incident sounds by the pinnae, or external ear. The ears function as a computer-steerable array similar to an electronically swept radar antenna. The form of transfor-

(Over)
2 cards, 4 copies

U.S. Naval Ordnance Test Station
Localization of Sound, Part 3. A New Theory of Human Audition, by United Research, Inc. China Lake, Calif., NOTS, December 1963. 28 pp. (NOTS TP 3109, Part 3), UNCLASSIFIED.

ABSTRACT. Localization of sound by humans has been shown to depend on a transformation of incident sounds by the pinnae, or external ear. The ears function as a computer-steerable array similar to an electronically swept radar antenna. The form of transfor-

(Over)
2 cards, 4 copies

NOTS TP 3109, Part 3

mation is that of time delays. Autocorrelation of the time delays by mental function provides localization.

It has been found that the ability to localize sounds in another environment may be reproduced using microphones adapted with ear replicas and high-quality condenser headphones. Extension of this technique to underwater use has been effectively demonstrated, despite some component shortcomings. The experimental highlights that support the theory and the field tests to evaluate the operational utility of localizing systems are discussed.

(Contd. on Card 2)

NOTS TP 3109, Part 3

mation is that of time delays. Autocorrelation of the time delays by mental function provides localization.

It has been found that the ability to localize sounds in another environment may be reproduced using microphones adapted with ear replicas and high-quality condenser headphones. Extension of this technique to underwater use has been effectively demonstrated, despite some component shortcomings. The experimental highlights that support the theory and the field tests to evaluate the operational utility of localizing systems are discussed.

(Contd. on Card 2)

NOTS TP 3109, Part 3

mation is that of time delays. Autocorrelation of the time delays by mental function provides localization.

It has been found that the ability to localize sounds in another environment may be reproduced using microphones adapted with ear replicas and high-quality condenser headphones. Extension of this technique to underwater use had been effectively demonstrated, despite some component shortcomings. The experimental highlights that support the theory and the field tests to evaluate the operational utility of localizing systems are discussed.

(Contd. on Card 2)

NOTS TP 3109, Part 3

mation is that of time delays. Autocorrelation of the time delays by mental function provides localization.

It has been found that the ability to localize sounds in another environment may be reproduced using microphones adapted with ear replicas and high-quality condenser headphones. Extension of this technique to underwater use had been effectively demonstrated, despite some component shortcomings. The experimental highlights that support the theory and the field tests to evaluate the operational utility of localizing systems are discussed.

(Contd. on Card 2)

ABSTRACT CARD

U. S. Naval Ordnance Test Station

<u>Localization of Sound</u> . . . (Card 2)

The basic concept of autocorrelation of time delays introduced by the pinna has been extended to speech recognition problems. A new theory of human audition, which ascribes significance to the time domain rather than the frequency domain, has been developed that explains not only binaural and monaural localization, but also the "cocktail party effect," pitch discrimination, speech recognition, masking, intelligibility in reverberation, and other auditory phenomena.

NOTS TP 3109, Part 3

U. S. Naval Ordnance Test Station

Localization of Sound . . . (Card 2)

The basic concept of autocorrelation of time delays introduced by the pinna has been extended to speech recognition problems. A new theory of human audition, which ascribes significance to the time domain rather than the frequency domain, has been developed that explains not only binaural and monaural localization, but also the "cocktail party effect," pitch discrimination, speech recognition, masking, intelligibility in reverberation, and other auditory phenomena.

NOTS TP 3109, Part 3

U. S. Naval Ordnance Test Station

Localization of Sound . . . (Card 2)

The basic concept of autocorrelation of time delays introduced by the pinna has been extended to speech recognition problems. A new theory of human audition, which ascribes significance to the time domain rather than the frequency domain, has been developed that explains not only binaural and monaural localization, but also the "cocktail party effect," pitch discrimination, speech recognition, masking, intelligibility in reverberation, and other auditory phenomena.

NOTS TP 3109, Part 3

U. S. Naval Ordnance Test Station

Localization of Sound . . . (Card 2)

The basic concept of autocorrelation of time delays introduced by the pinna has been extended to speech recognition problems. A new theory of human audition, which ascribes significance to the time domain rather than the frequency domain, has been developed that explains not only binaural and monaural localization, but also the "cocktail party effect," pitch discrimination, speech recognition, masking, intelligibility in reverberation, and other auditory phenomena.

NOTS TP 3109, Part 3

INITIAL DISTRIBUTION

6 Chief, Bureau of Naval Weapons
  DLI-31 (2)
  RRRE (1)
  RU (1)
  RUTO (1)
  RUTO-2 (1)
3 Special Projects
  SP 20 (1)
  SP 22 (1)
  SP 27 (1)
1 Chief, Bureau of Ships (Code 560)
1 Chief of Naval Operations
2 Chief of Naval Research
  Code 104 (1)
  Code 466 (1)
1 David W. Taylor Model Basin
1 Fleet Anti-Air Warfare Training Center, San Diego (Guided Missile Section)
1 Naval Advanced Undersea Weapons School, Key West
1 Naval Air Development Center, Johnsville
1 Naval Engineering Experiment Station, Annapolis
1 Naval Ordnance Laboratory, Corona
1 Naval Ordnance Laboratory, White Oak
1 Naval Postgraduate School, Monterey (Library, Technical Reports Section)
1 Naval Research Laboratory
1 Naval Torpedo Station, Keyport (Quality Evaluation Laboratory, Technical Library)
1 Naval Underwater Ordnance Station, Newport
1 Navy Central Torpedo Office, Newport
2 Navy Electronics Laboratory, San Diego
1 Navy Mine Defense Laboratory, Panama City
1 Navy Underwater Sound Laboratory, Fort Trumbull
1 Navy Underwater Sound Reference Laboratory, Orlando
1 Norfolk Naval Shipyard (Underwater Explosion Research Division)
1 Atlantic Missile Range
1 Pacific Missile Range, Point Mugu
1 Submarine Development Group 2
10 Defense Documentation Center (TISIA-1)
1 National Aeronautics & Space Administration
1 Acoustics Laboratory, Harvard University, Cambridge
1 Aerojet-General Corporation, Azusa, Calif., via BuWepsRep