A STUDY OF ACOUSTICAL MULTIPATH SYSTEMS

SECTION I

Theory and Program

by

Dwight W. Batteau

1 June 1966

Final Report: Contract NONR 494(00)

Prepared for

U. S. Navy Office of Naval Research

Department of Mechanical Engineering
TUFTS UNIVERSITY
Medford, Massachusetts

1.1 Introduction

The investigation of the effects of multiple paths on acoustical signals as related to speech recognition and improvement of signal-to-noise ratios has involved the following:

1. Selection of talker and preparation of phonetically balanced (PB) word lists.
2. Development of computer programs to handle the data on IBM 1620.
3. Construction of semi-insertion capacitor headphones. (None available commercially.)
4. Construction of ear molds and casts to provide a source of pinnae.
5. Construction of pinna and microphone mounting to provide careful control of the acoustic path.
6. Construction of test apparatus for the headphones.
7. Conduct of a systematic investigation of the effects of variations in headphones on subjective speech intelligibility.
8. Preliminary tests of the pinna pick-up system.
9. Acquisition of needed tape recording equipment.
10. Construction of control room equipment.
11. Construction of summing amplifiers for adding multiple pickups.
12. Adapting the nickel wire acoustical delay line to analysis and synthesis of multipath single channel signals.
13. Construction of multiple microphone test installation.
14. Preliminary tests of multipath synthesis in the pinna domain using N delay lines.
15. Continued theoretical development in association with other programs.

1.2 Basic Theory

The basic theoretical structure being applied to the effects of multiple acoustical paths is derived principally from information theory and its associated field of transformations as encodings. Every iteration of a signal by an additional path is a redundance in the message, and may be considered to be a secondary encoding. If we consider speech to be a consequence of transformation of a vocal pulse, we may define the following applicable transformations.

$D(s)$ = a Dirac pulse of initiation.

$T_1(s) \triangleq$ the transformation of the initiating Dirac into a vocal pulse.

$T_k(s) \triangleq$ the set of transformations of the vocal pulse by the vocal tract.

$T_n(s) \triangleq$ the set of environmental transformations applicable to a variety of sound source locations, n , in a room.

${}_rT_n(s) \triangleq$ the set of sets of location of sound sources applicable to different rooms, r.

The creation of vocal pulse may be taken as a resultant of transformation of the Dirac of initiation by a linear active system containing energy sources. We write equation (1.2.1)

$$P_v(s) = D(s)\, T_1(s) \tag{1.2.1}$$

Our assumption that $T_1(s)$ is an active linear transformation and that $D(s)$ is an initiation contributing no energy to the result permits us to place a restraint on $P_v(s)$ and $T_1(s)$ by a measure of energy. We may, knowing it to be an assumption, assign the energy to the Dirac and consider $T_1(s)$ to be a linear passive transformation which redistributes energy in time, space and form. An inverse to the transformation $T_1(s)$ then recovers the Dirac, concentrating energy in a single, sharp pulse. If we consider the set of interest to be all quanta of energy which are redistributed by the transformation, given a need to construct the inverse, not knowing $T_1(s)$, we may proceed to fashion $R_1(s)$ to concentrate the distribution back into a single pulse.

If

$$P_v(s)\, R_1(s) = D(s) \tag{1.2.2}$$

then

$$R_1(s) = T_1^{-1}(s) \tag{1.2.3}$$

We also may accept as a measure of fit of $R_1(s)$ to the proper $T_1^{-1}(s)$ the degree of concentration. If

$$\overset{\frown}{R_1(s)\, P_v(s)} = \text{peak value of energy}$$

$$\overline{R_1(s)\, P_v(s)} = \text{total value of energy}$$

then

$$\frac{\overset{\frown}{R_1(s)\, P_v(s)}}{\overline{R_1(s) P_v(s)}} = M(R_1) \tag{1.2.4}$$

$$0 \leq M(R_1) \leq 1$$

When

$$M(R_1) = 1$$

the Dirac is reconstructed. We may assign M(R) to the class of probabilities.

Let us assume that

$$P_v(s) = D(s)\, T_1(s)$$

is stationary for a given example of speech, as for a given person in a given time span. Then speech involves a time sequence of the form implied by equation (1.2.5)

$$H_k(s) = P_v(s)\, T_k(s) \tag{1.2.5}$$

If the range of k exhausts the class of speech transforms, a particular spoken message consists of a time sequence of reorderings of the values of k. For example, let

$k = 4 \longrightarrow$ ah

$k = 7 \longrightarrow$ eeh

$4, 7 \longrightarrow$ aheeh

$7, 4 \longrightarrow$ eehah

Every spoken sequence is then an ordering of values of k, which in a linear space may be thought of as a vector formed from a basis attributable to $T_k(s)$. It is clear, thus, that not only the final vector, but the ordering of base elements in its construction are significant, and in fact, the order carries the message. Speech recognition then requires (it may not be sufficient) that the sequence of k's be constructed.

Let us assume that the k sequence is generated in a room, or any reverberant environment, such that the transformation may be written

$$ {}_nH_k(s) = T_k(s)\, T_n(s) \tag{1.2.6} $$

for a single value of n and a sequence of k. It is now not clear in the resultant which factors are attributable to k and which to n, unless $T_n(s)$ may be known.

We may, however, consider that $T_n(s)$ redistributes each $T_k(s)$ in a stationary sequence characteristic of the reverberation. So, for a particular k and a given n, we may write for reverberant multipath,

$$ {}_nH_k(s) = T_k(s) \sum_{i=1}^{\infty} a_i \epsilon^{-s\tau_i} \tag{1.2.7} $$

Thus, any single sample of the redistribution would be attenuated compared to the original, but if $T_n^{-1}(s)$ could be found for the given condition, the reverberant iterations of $T_k(s)$ could be brought together coherently and thus energy concentrated in a single expression. It is conceivable that signal-to-noise ratios could be operationally improved if the given $T_n(s)$ did not apply to any of the noise sources.

If we apply the same reasoning to consider the effect of pinna as local reverberation systems, or multipath structures, application of the

appropriate transformation inverse to that for a particular locale also concentrates energy coherently only in the signal from that point.

Finally, any multipath system may be considered to introduce redundances of the basic message, and it is conceivable that these may be used to select a particular message source from among many by making use of the characteristic transformation between the source and observer applicable to the desired relationship. We have called transformations which do this "attention functions".

1.3 Operational Considerations

If we wish to make use of the multipath transformations, we can begin with consideration of the consequence of manipulation as represented mathematically. Let us consider first the generation of speech.

$$H_k(s) = P_v(s)\, T_k(s)$$

$$= D(s)\, T_1(s)\, T_k(s) \qquad (1.3.1)$$

And let us assume that the individual has a characteristic $T_1(s)$, and write

$$W_k(s) = H_k(s)\, T_1^{-1}(s)$$

$$= D(s)\, T_k(s) \qquad (1.3.2)$$

This implies that the energy distributed into the vocal pulse structure can be concentrated in the statement of the speech transformation. If individual (personal) vocal pulses are distinguishable transformations, these can be used to provide attention functions. As an approach to derivation $T_1(s)$ may be invariant in the set $T_k(s)\, T_1(s)$, and thus a continued examination of the formed construction of

$$[T_k(s)\, T_1(s)]^{-1} \qquad (1.3.3)$$

examined for such an invariance. This may be paraphrased as "learning to recognize a voice". (I expect that other factors also contribute to personal voice character.)

Similarly, we may assume that a person speaking is standing still, and that a particular room transformation is invariant in the set ${}_nH_k(s)$, for a fixed value of n. This would permit search for invariance in the construction of

$$[T_k(s)\ T_n(s)]^{-1} \tag{1.3.4}$$

In operation, we would thus construct, by the measure given, an inverse to a signal, then seek to find partition in the signal by means of invariances in subsets of the signal for which inverses are also constructed. If we allow separable time domains of iterations (not necessary but convenient), we would find duplications of sequences in the various delay segments of the sequences of exact attention functions. This would be applicable to a single voice in a reverberant room to catch the room and vocal pulse characteristics, the room delays being long, the vocal pulse being short. A period of "quiet" is then useful to provide "capture" for "tracking".

1.4 Program

We have assumed that the first available computer having suitable programming facilities for the utilization of multipath acoustical signals is the human mind. (We are also pursuing machine computation on a 7090.) Thus we have designed our study to examine first the effects on subjective speech intelligibility of a variety of multipath systems which are to include pinna, synthetic pinna, multiple microphones, room reverberation, and a variety of synthetic and processed signals. In general, we have assumed that listening through headphones is desirable to remove environmental effects at that point. We have also assumed that human speech is

the signal of interest, although we plan eventually to use synthetic signals of measurable equivalence to human speech.

Thus, our early phases were concerned with getting truly high fidelity tapes (20 Hz to 30 KHz) tapes of human speech, most commercially available recordings being limited to the hearing aid range (300 Hz to 5KHz). We also concerned ourselves with the quality of headphones which would preserve the wide bandwidth we wish to have in our investigation. And we concerned ourselves with the design and construction of a pickup with pinnae to transfer human hearing as naturally as possible.

The choice of talker and the making of tapes was completed in September 1965, the useful headphones in December 1965, and the pinna pickup almost completed by May 1966. Even though all was not perfect, preliminary tests with talker tapes, headphones, and pickup have been conducted. We are now at a testing point for several systems (pinna, multiple microphones, delay lines).

We plan to improve the talker tapes and studio equipment, employ the pinna pickup (Marc Anthony I or Marc I), the multiple microphones, and the delay lines in studies during the summer 1966. In the fall 1966, we plan acoustical complex and room reverberation studies, synthetic speech of proper measure, localization synthesis and delay line and computer processing.