1968-06

An automatic electronic speech teaching responder.

Kuligowski, Theodore Joseph

Monterey, California. U.S. Naval Postgraduate School

AN AUTOMATIC ELECTRONIC SPEECH TEACHING RESPONDER

by

Theodore Joseph Kuligowski

UNITED STATES
NAVAL POSTGRADUATE SCHOOL

THESIS

AN AUTOMATIC ELECTRONIC SPEECH TEACHING RESPONDER

by

Theodore Joseph Kuligowski

June 1968

~~This document is subject to special export controls and each transmittal to foreign government or foreign nationals may be made only with prior approval of the U. S. Naval Postgraduate School.~~

AN AUTOMATIC ELECTRONIC

SPEECH TEACHING RESPONDER

by

Theodore Joseph Kuligowski
Lieutenant Commander, United States Navy
B.S., Villanova University, 1955

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
June, 1968

ABSTRACT

A concept for an apparatus which visually displays and responds to the first and second formant of vowel sounds is developed. The machine is intended for use by deaf and speech handicapped children in learning to produce voiced sounds. System design and principles applied to realize a physical prototype of this concept are presented. The complete electronic and mechanical design plus fabrication of the automatic electronic speech training responder is described in detail. Schematic diagrams of all electronic circuitry employed and photographs of the prototype equipment are included. The apparatus is on loan to the Monterey Institute for Speech and Hearing, Monterey, California, for clinical testing and evaluation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

SYMBOLS AND ABBREVIATIONS

F1	First formant frequency in Hertz
F2	Second formant frequency in Hertz
I_D	DC drain current
I_{ds}	Total instantaneous drain to source current
I_{DSS}	Common source zero gate voltage drain current
R_D	Drain resistor between FET drain and power supply
R_s	Self bias resistance
V_{DD}	Drain supply voltage
V_{GS}	DC gate to sour voltage
V_{gs}	Total instantaneous gate to source voltage
V_o	Oscillator voltage, peak value
V_P	Pinch off voltage
V_s	Signal voltage, peak value

ACKNOWLEDGEMENTS

The author is deeply indebted to Dr. Gerald D. Ewing of the Naval Postgraduate School faculty for arousing interest in the subject of electronic speech aids through the 1967 edition of the IEEE Student Branch Thesis Topics publication. His guidance and encouragement throughout the development of the prototype machine, in large measure, made it possible to have an apparatus which can be tested in a clinical environment.

The author also expresses his appreciation to Dr. Burl Gray of the Monterey Institute for Speech and Hearing for his valuable assistance and time in defining problem areas of speech research plus stating criteria electronic devices must meet to serve the needs of speech handicapped children. His many helpful suggestions ensured that the machine's final objective did not become distorted during the circuit design and fabrication stages.

Many members of the Naval Postgraduate School electronics laboratory provided assistance in the design and construction of AESTR. Mr. Victor McCullough fabricated and wired the chasis, switches and a portion of the circuit boards. T. Kedzie, ETSN applied his artistic talents to lay out and letter the final version of the AESTR control panel. Mr. Walter Landaker and Mr. Hollis W. Oren provided much valuable advice concerning the practical aspects of circuit design and fabrication. Both gentlemen also made available the necessary integrated circuit devices and other components necessary for the fabrication of AESTR.

A special note of thanks is extended to Mrs. C. LaVerne Cooper and Mrs. Nancy I. Hill of the Defense Language Institute Academic Library who kept the author informed during the past year of all publications related to speech analysis and research.

[illegible]

1. INTRODUCTION

Man is born with the natural instinct and physical capacity to eat and breathe, but he must <u>learn</u> how to speak. This learning process depends on good hearing ability during the formative years. A child, during initial attempts to speak, constantly monitors his utterances with his ears. These sensors provide the necessary information to the brain to modify the vocal tract modulators and articulators with respect to the points of articulation until the desired sound, phoneme or word is correctly produced. If this feedback loop (voice output-ear sensor-brain input) is defective or nonexistent in a human, it is necessary that another physical sensor must be used as an alternate feedback path to monitor generated speech sounds on a real time basis if intelligent and comprehensible communication is to be achieved. Many devices have been devised and constructed which transform speech sounds into a visual display or a tactile signal.

This thesis is directed toward the attempt to process specific speech sounds and to display or provide a positive response when the desired sound has been correctly produced. In addition, the machine must be simple in final output so that it can be easily used and interpreted by children.

Computer sciences have stimulated research into speech recognition and synthesization. Unfortunately, this type of engineering technology is too costly and sophisticated at the present time for application to elementary speech training problems. Rather specific guide lines on needs of training devices for children were developed by Dr. Burl Gray of the Monterey Institute for Speech and Hearing; these are:

1. A definite need exists for simple, inexpensive devices which will assist or supplement the speech therapist's work with deaf children. These devices would permit the instructor to teach more students simultaneously or the devices could perform elementary tasks of providing various mechanical responses to repetitious articulation drills without the constant attention or intervention of the speech therapist.
2. The information display or mechanical response of such an apparatus must be in a form which is easily communicable to and understood by the child. Careful attention must be given to the human-machine interface problem to insure good results with a given age group and mental attitude.
3. The apparatus must present the visual or mechanical response while the child is speaking (i.e. real time).

Using these criteria, an attempt has been made to design and construct an apparatus which will respond only to a defined pronounciation of the basic American vowel sounds. The vowel sounds were selected for machine recognition because they require the minimum amount of audio spectral information to be uniquely identified. However, the approach to this vowel processing technique is sufficiently general, It may have possible extensions to process other sounds.

Figure 1 is a graphic representation of a generalized man-machine speech feedback system.

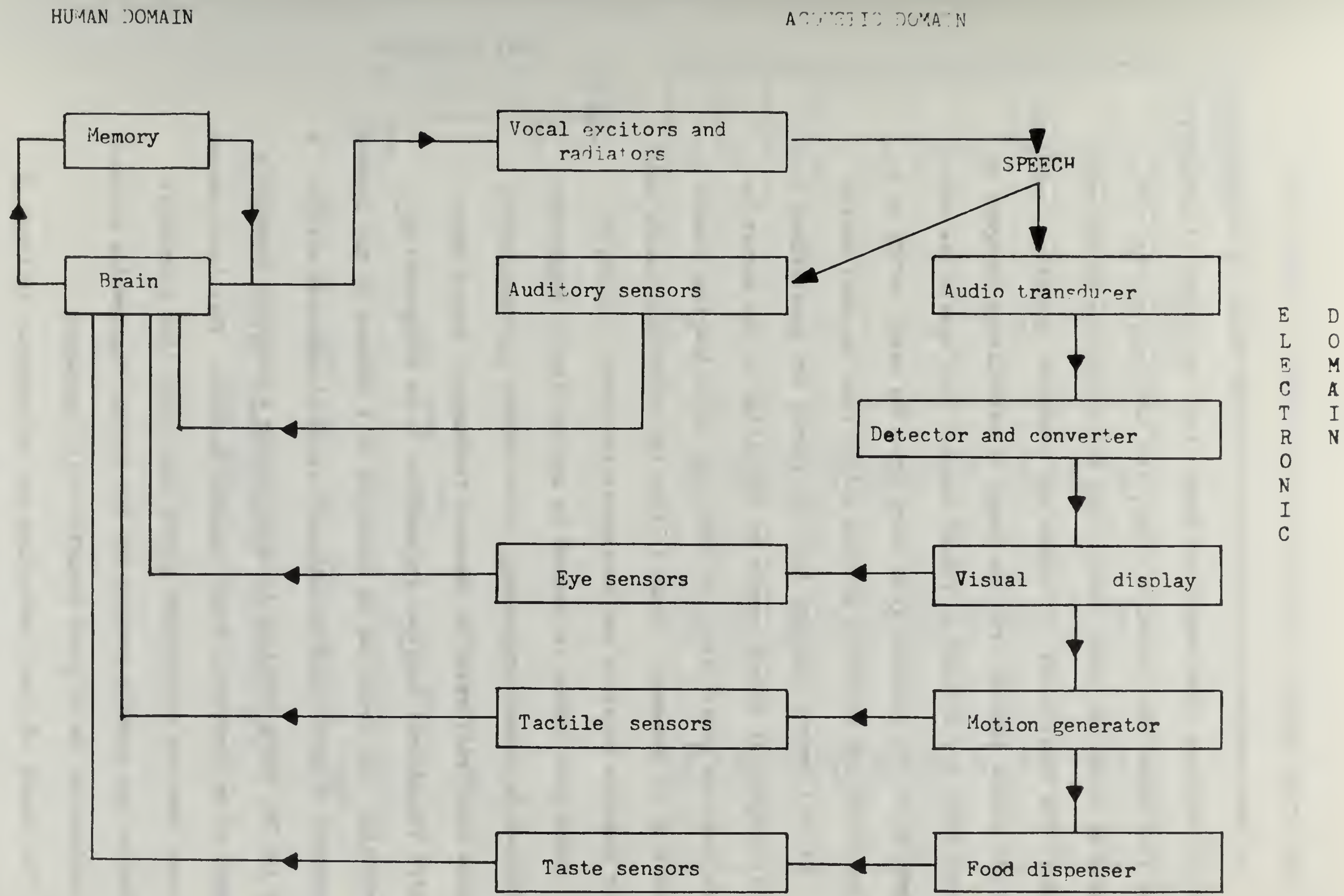


Figure 1 - Man-Machine Speech Feedback System

2. THE VOWEL SOUND

A human can produce a multitude of speech sounds by controlling his articulators (the tongue and lower lip), the points of articulation (the upper lip, alveolar ridge, the hard palate, the soft palate or velum and lower teeth), and the excitation of his vocal bands. The vocal bands, if tensed and therefore vibrating, modulate the air stream exhaled from the lungs to establish a category of sounds which are classed as voiced sounds. All vowels are voiced sounds which are excluded from entry into the nasal cavity by a raised velum and therefore eminate solely from the oral cavity.

It will become apparant that the vowels are constrained to a small category of speech sounds by definition of the manner in which they are articulated. In fact, the basic American vowels consist of 10 phonemes. Table 1 lists the individual sounds with their phonetic notation and representative words. [10,30,33]

Since this thesis is devoted to application of electronic techniques to speech processing, it is natural to begin with a machine which will react to the most fundamental sounds which require the minimal spectral information to be recognized or identified. The vowel can be specified by a minimum of two spectral parameters in most sound situations. Joint discussions with Dr. Gray and Dr. Ewing resulted in establishing a mutually acceptable concept of an electronic vowel teaching machine. This local merger of ideas from two disciplines proves once again that scientific boundaries can greatly overlap and the systems engineering approach to problems may be of great benefit to all concerned.

The theory of vowel production can be described in terms of steady state (or harmonic) conditions with application of

TABLE 1

VOWEL PHONETIC SYMBOLS AND REPRESENTATIVE WORDS

Typewritten Symbol for Vowel	IPA Symbol	Representative Words		
IY	i	heed	beat	eat
I	ɪ	hid	bit	it
E	ɛ	head	bet	let
AE	æ	had	bat	hat
A	a	hod	calm	father
OW	ɔ	hawed	fall	lost
U	ʊ	hood	full	foot
OO	u	who'd	fool	pool
UH	ʌ	hud	above	tub
ER	ɝ	heard	word	hurt

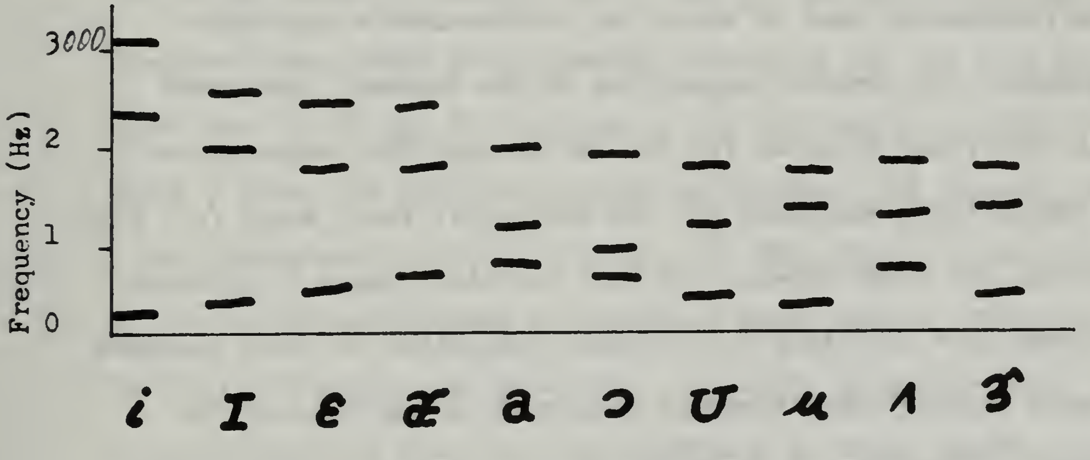


Figure 2. Typical Spectrograms of the vowels by a male voice.

cord-tone-resonance effects. [21] Modern analytical representation of the same effect can be stated in terms of excitation functions and convolution techniques. [36] The former description states in effect that the vocal bands (a modern synonym for cords [3]), during phonation, set up in the air immediately adjacent to them a complex motion which consists of a fundamental component, known as pitch, and a large number of its overtones or harmonics. This complex air motion constitutes the so-called band-tone. The theory further states that the vocal cavities, on which the band-tone acts as a force, have the properties of simple resonators and thus serve to modify the spectrum of energy flowing from the bands. In terms of this theory, a vowel sound, as emitted from the mouth, is due to both selective generation and selective transmission plus radiation. This sound is composed mainly of harmonic components of the fundamental each of which has a determinable magnitude. For example, the greatest magnitudes of the harmonic components usually are found to exist for the 6th through 9th component and 13th through 16th component for the particular vowel sound /a/. [21] Naturally, for other vowels, the oral cavities change in physical dimensions thus affecting the resonant properties of these chambers and hence causing other harmonic components or partials of the fundamental vibration of the vocal bands to be amplified or attenuated.

The spectograph has greatly enhanced the study of speech sounds and in particular vividly identifies the amplified partial tones or resonant frequencies uniquely identifiable with each vowel sound. [32,33] Figure 2 provides a sketch representing the spectrographic tracings due to each vowel sound. The dark areas represent the amplified harmonics of the fundamental pitch of the voice. Note that these locations are unique for each vowel, especially for the first and second resonant frequencies. In the terminology of visible speech the dark bands are called "formant" regions or "bars" and for reference purposes are designated by number, the lowest on the frequency scale being bar 1 or first formant F1, the next bar 2 or second formant F2, etc. In this thesis, the notation F1 and F2 shall be used to designate the first and second resonant frequencies of vowel sounds respectively.

The first and second formants are the only two pieces of spectral information required, in most cases, to identify a particular vowel. The third formant (F3) is helpful in distinguishing between overlapping first and second formant frequencies. Potter and Peterson have suggested that the human ear recognizes vowel sounds, not by the spectral location of F1 and F2, but rather by the relative frequency separation or difference between F1 and F2. [33] Table 2 lists the F1, F2 and F3 frequencies for the vowels of Table 1 while Table 3 lists the relative formant amplitudes. [30] Figure 3 shows a two dimensional plot of F1 vs F2. [9] This figure is the crux of the apparatus designed to recognize vowel sounds. Note that in the F1-F2 plane each vowel has a specific location; also it is interesting to note that the locations of these sounds corresponds roughly to the position of the tongue in the oral cavity if you imagine looking at a side view of the head.

The vowel training device does not work on the relative location of F1 to F2 but rather utilizes an electronic spectral window in the F1-F2 plane to target a particular vowel sound or for that matter, any voiced combination of two oral resonances in this dual formant plane.

TABLE 2

AVERAGES OF FUNDAMENTAL AND FORMANT FREQUENCIES

IPA Symbol		Fundamental Frequency (Hz)	First Formant(Hz)	Second Formant(Hz)	Third Formant(Hz)
i	M	136	270	2290	3010
	W	235	310	2770	3310
	Ch	272	370	3200	3730
I	M	135	390	1990	2550
	W	232	430	2480	3070
	Ch	269	530	2730	3600
ɛ	M	130	530	1840	2480
	W	223	610	2330	2990
	Ch	260	690	2610	3570
æ	M	127	660	1720	2410
	W	210	860	2050	2850
	Ch	251	1010	2320	3320
ɑ	M	124	730	1090	2440
	W	212	850	1220	2810
	Ch	256	1030	1370	3170
ɔ	M	129	570	840	2410
	W	216	590	920	2710
	Ch	263	680	1060	3180
ʊ	M	137	440	1020	2240
	W	232	470	1160	2680
	Ch	276	560	1410	3310
u	M	141	300	870	2240
	W	231	370	950	2670
	Ch	274	430	1170	3260
ʌ	M	130	640	1190	2390
	W	221	760	1400	2780
	Ch	261	850	1590	3360
ɝ	M	133	490	1350	1690
	W	218	500	1640	1960
	Ch	261	560	1820	2160

TABLE 3

FORMANT AMPLITUDES MEASURED RELATIVE TO /ɔ/

IPA Symbol	First Formant (db)	Second Formant (db)	Third Formant (db)
i	-4	-24	-28
ɪ	-3	-23	-27
ɛ	-2	-17	-24
æ	-1	-12	-22
ɑ	-1	-5	-28
ɔ	0	-7	-34
ʊ	-1	-12	-34
u	-3	-19	-43
ʌ	-1	-10	-27
ɝ	-5	-15	-20

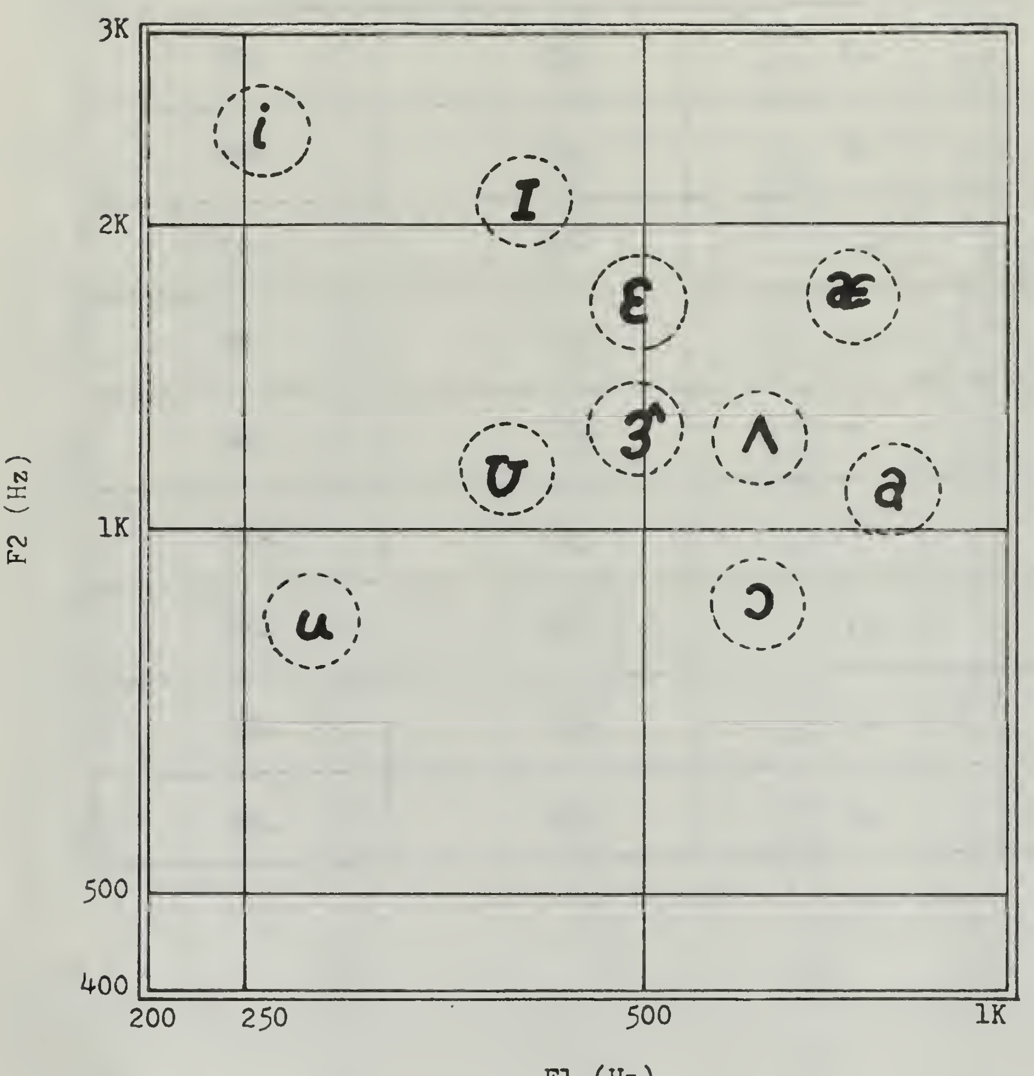


Figure 3. Central Regions of First and Second Formant Frequencies of the Common American Vowels.

3. AESTR AS A TEACHING AID

During the process of physically realizing a prototype of the vowel training machine, liberty was taken by the author and his associates in the electronics laboratory to coin a name for this device. The result is Automatic Electronic Speech Teaching Responder. The title should convey the notion that this machine is not intended to replace a speech therapist but rather assist him in his work. AESTR will be initially preset through the oscillator frequency dials and the control knobs located on the front panel of the device. Now the child is placed in a room with a candy dispenser or some other motivational responder and a microphone. He is asked to make any sound he cares to. As the child produces various sounds he should produce the desired sound in due time. The machine will only activate the candy dispenser when the child has produced the targeted voiced sound and the child will keep trying to repeat the sound in order to maximize his reward. As the rewards are given more frequently, the teacher is able to adjust the filter bandwidths on AESTR and narrow the spectral window of the desired sound, hence increasing the articulation accuracy required of the child if he is to obtain his reward.

The child learns to speak desired sounds by communicating directly with AESTR. However, positive control of the speech training process is available to the teacher by his ability to vary six parameters from the front panel of AESTR. (F1 bandwidth and sensitivity, F2 bandwidth and sensitivity, pitch filter, microphone gain).

4. SYSTEM CRITERIA AND DESIGN

The system incorporates two basic electronic functions to locate and measure, in real time, the first and second formants of a voiced sound. The speech sound is first mixed with two local oscillators by means of non-linear devices. One of the components obtained from this process, the difference frequencies between the formants and oscillators, is isolated by active low pass filter circuits. The oscillators and low pass filters are variable and can be set for a particular sound or spectral window. By setting the two local oscillators to the known frequencies for F1 and F2 of a particular vowel and the low pass cutoff frequency for the desired degree of accuracy of response, the machine is able to process the speech sound and provide a binary decision response.

The responses are:

1. A positive response which is movement of two voltmeter indicators and a light being activated if both meters are at maximum value simultaneously. This condition occurs when the resonant frequencies of the voice correlate with the preset local oscillator frequencies simultaneously. The correct voiced sound is being produced by the student. The apparatus also has an external motivation output jack which can operate other reward machines when the targeted sound is produced by the student.
2. No response. One or both formant frequencies are not present or they do not correlate within limits set by the filter pass band.

Many methods were considered for realization of this device in terms of simplicity, cost and expediency. Primary concern was to produce some type of primitive machine which would do the basic tasks required by this particular vowel teaching aid. The approach finally selected for the first attempt is to process the complex speech waveform in analog form in the audio specturm. Advantage was taken of the Field-Effect Transistor(FET) which has an almost perfect square law response and which is ideally suited for optimum mixing of oscillator and voice frequencies. The filtering is accomplished by means of active low-pass filters using the readily available integrated circuit operational amplifiers.

An additional factor must be considered in AESTR's system design. The pitch of a human voice can range from approximately 75 to 500 Hz. [32] The formant frequencies range from approximately 250 to 3000 Hz. It is necessary to eliminate the pitch frequency from the audio speech prior to the mixing operation, otherwise it is possible for the pitch or fundamental frequency to pass directly through the mixers and filters thus producing a positive machine response regardless of the formant and oscillator frequencies present. Figure 4 represents the basic system approach for realization of this apparatus.

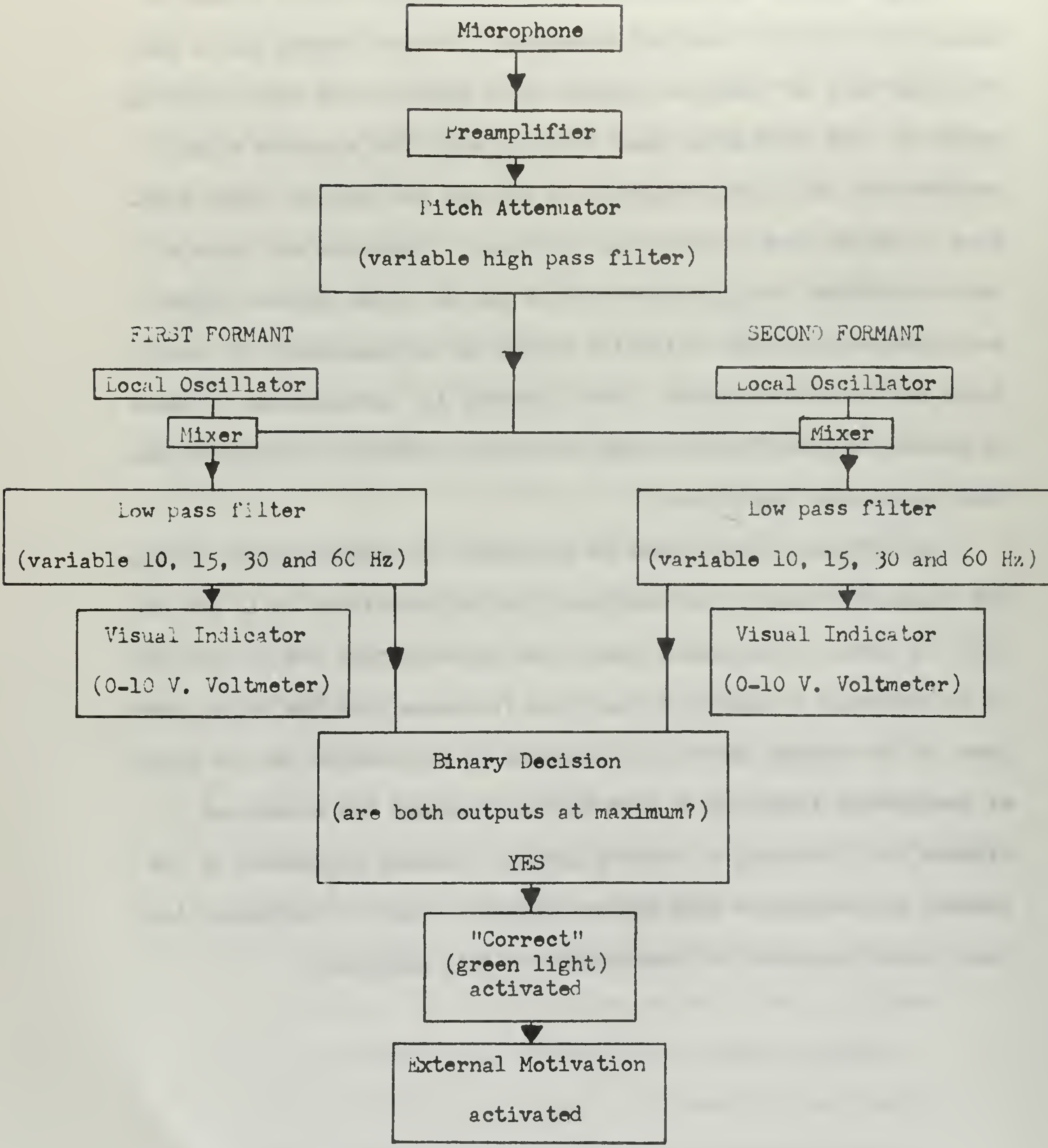


Figure 4. AESTR Electronic Transducer, Detector, and Display System

5. CONTROL PANEL DESIGN AND OPERATION

AESTR is to be operated by individuals who do not possess an engineering background. Therefore the panel is designed to be self explanitory and requires minimum instruction for operation. The controls are fairly large to permit positive grasp by the operator. Also functional location of the knobs and visual indicators is evident by the partition lines. The objective is to have the panel functions reflect the needs of the operator rather than the requirements of the internal circuitry. AESTR's control panel is shown in Figure 5.

The "volume" control is self explanitory and permits the operator to vary the gain of the preamplifier circuit.

The "pitch" control permits selection of four cutoff frequencies of the high-pass filter circuit in order to suppress the fundamental frequency of a voice while passing all the formant frequencies. In Table 4 below, the letter positions are identified with the 3 db cutoff frequencies of the high-pass filter.

TABLE 4

HIGH-PASS FILTER CUTOFF FREQUENCIES

Pitch control position	Frequency (Hz)
A (male voice)	75
B (female voice)	190
C (child's voice)	450
D (special use)	1050

The pitch control setting is not critical for the back vowel sounds such as OW in the word "father" and can remain in the "A" or "B" setting for all speakers regardless of sex or age.

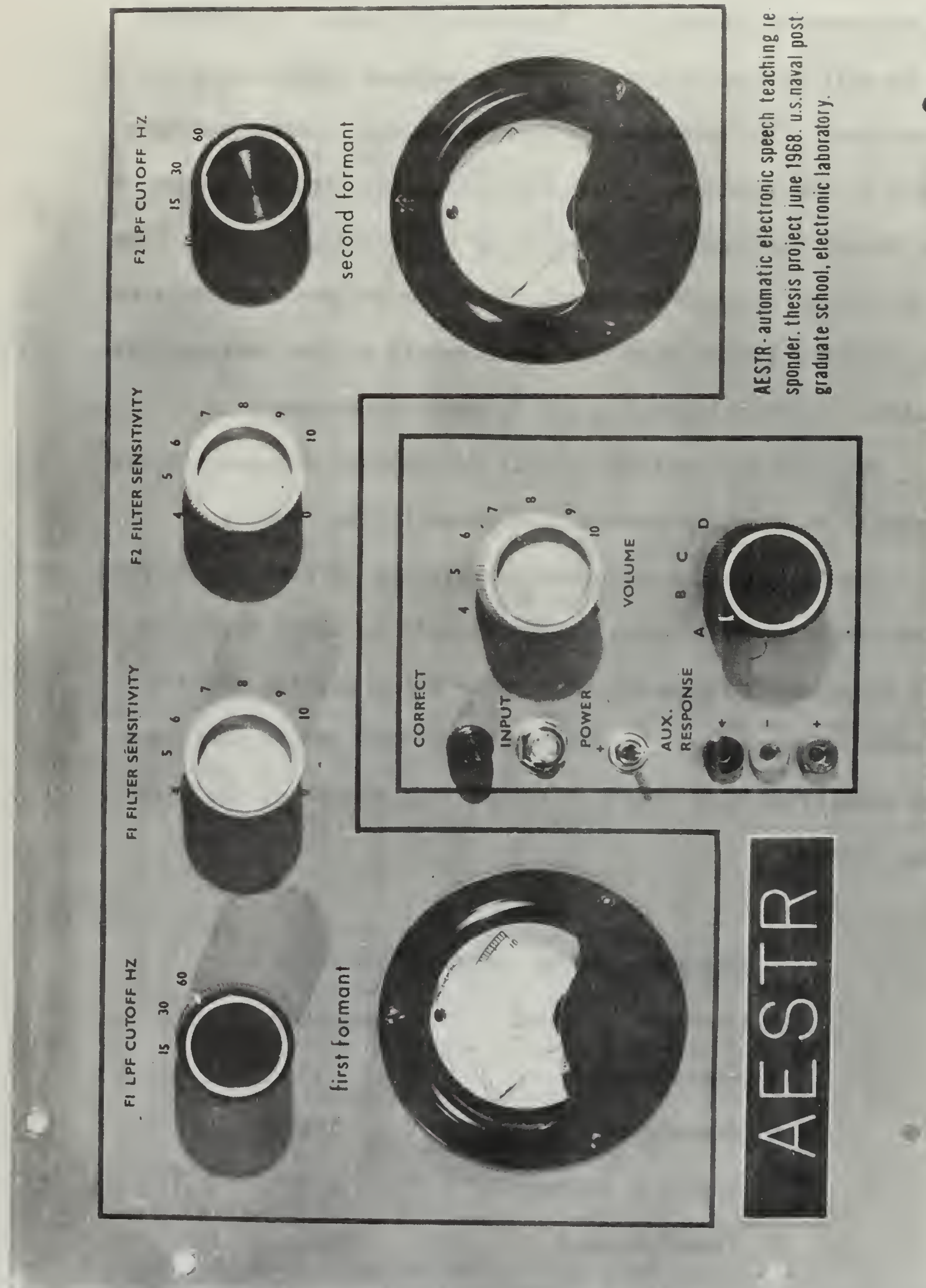


Figure 5. AESTR Control Panel

Position "D" is used when working with the central vowel such as ER in the word "bird". It is necessary to suppress the frequencies below 1 KHz and operate with the second and/or third formants in order to have the machine only respond to this particular vowel. This technique was developed during the testing of AESTR and is discussed further in section 13.

The "F1 LPF cutoff" and "F2 LPF cutoff" are variable cutoff low-pass filters. The controls are located above the first and second formant voltmeter indicators respectively. Cutoff frequencies of 10, 15, 30 and 60 Hz are printed around the periphery of the control knobs. Normally the controls are initially set in the 60 Hz position when searching for voice formant. This setting provides the widest possible filter pass band, such that the voltmeter needle will begin to deflect up scale whenever the oscillator and voice formant are within ± 60 Hz of each other. As the two frequencies become more nearly coincident, the voltmeter needle will show a maximum scale deflection. When the operator has the oscillator set at a frequency which gives the maximum needle deflection, he may elect to switch the "LPF cutoff" control to 30 Hz in order to narrow the filter response pass band. It may be necessary to readjust the local oscillator slightly for maximum scale deflection. This procedure can be continued for the 15 and 10 Hz cutoff frequencies respectively.

The "F1 Filter Sensitivity" and "F2 Filter Sensitivity" controls vary the gain of the filters. The word "sensitivity" is chosen for contrast against the "volume" control nomenclature and is intended to prevent any misunderstanding between the two types of controls. The "filter sensitivity" controls are adjusted to make the voltmeter needle deflect to full scale when the local oscillator and formant frequencies most nearly coincide. Each vowel sound will have its unique "filter sensitivity" setting due to the varying intensity levels of the formants of the individual phonemes. The operator must determine these settings empirically since the sensitivity is also a function of the intensity of the speaker's voice. It is advisable to keep the "volume" control knob at a minimum setting and the "sensitivity" control knobs at a high setting to reduce the effects of acoustic and electrical noise.

The "correct" green light illuminates when both formant indicators read an up scale deflection of 7 volts. Light activation is delayed 250 milliseconds and once lighted, stays on for a period of 2 seconds. The delay prevents the light from being activated by transient full scale deflections which occur from plosive type consonant sounds preceding a vowel in such a word as "bar". The light hold time of 2 seconds prevents the light from flickering if the voice begins to quiver during articulation of a phoneme.

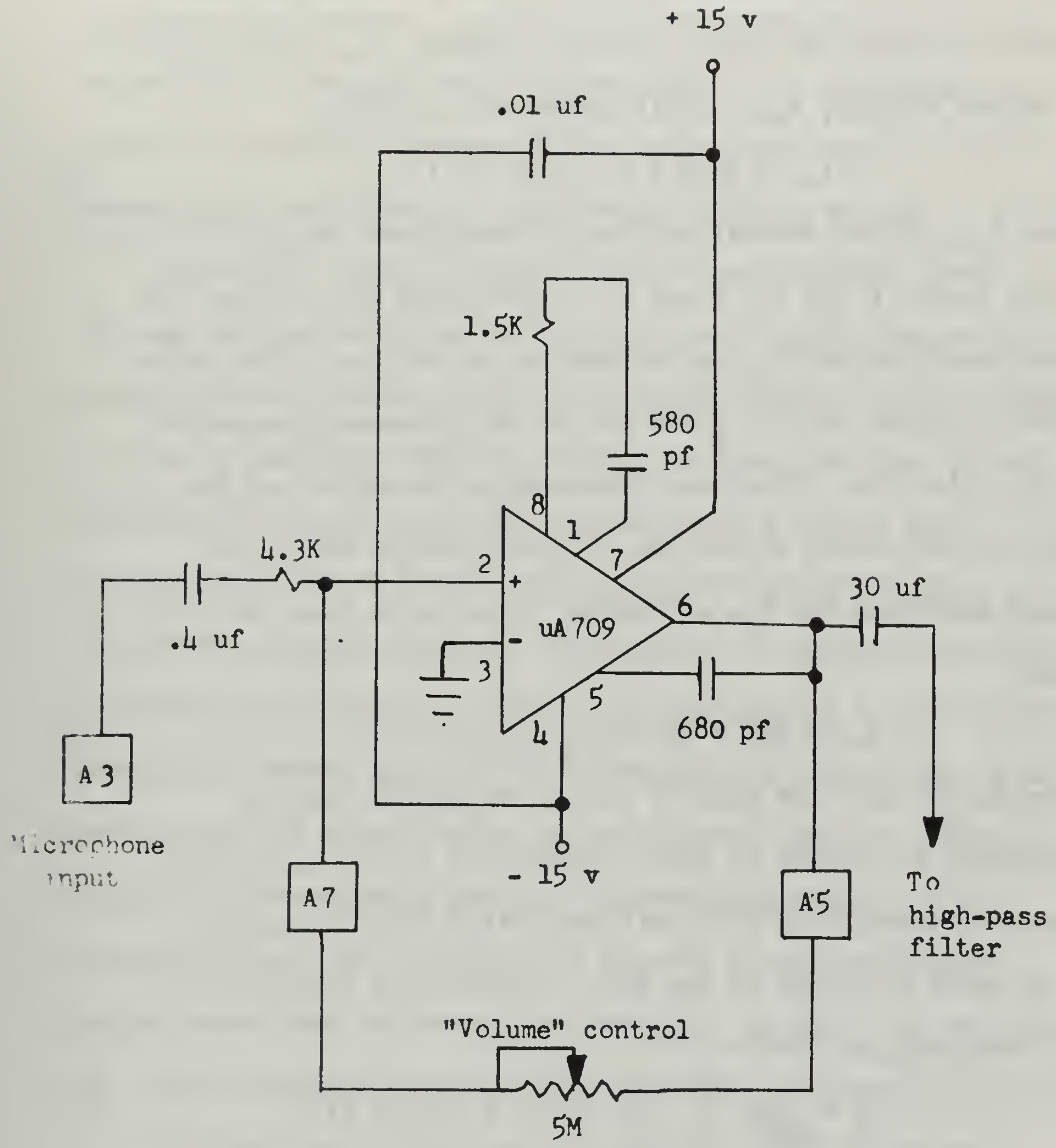
In the rear of the AESTR cabinet is located an ordinary female 115 volt receptacle. Any external motivational device, such as an M&M candy dispenser, can be attached to this terminal and will be operated automatically since the terminal provides 115 volts only during the interval when the "correct" light is illuminated.

AESTR also has the capability of measuring the pitch of a person's voice. Turn the "pitch" control clockwise beyond the "0" position until it stops. Set the "F1 LPF Filter" control to 10 Hz and the "F1 Sensitivity" control to a maximum value of 10. Turn the "F2 Sensitivity" full counterclockwise to a value of 0. Sweep the F1 local oscillator through a range of 60 to 500 Hz. The speaker's pitch will be read on the F1 oscillator frequency setting when the first formant visual indicator has a maximum up-scale deflection.

6. PREAMPLIFIER DESIGN

The mixer circuit is able to accept a maximum input signal of 0.8 volts peak to peak. A preamplifier is necessary, especially if a dynamic microphone is being used, to amplify the voice sound for maximum mixer output. The Fairchild uA709 operational amplifier was selected to perform this function untilizing the standard feedback configuration and necessary frquency compensation. It is shown schematically in Figure 6.

The uA709 comes in an epoxy TO-5 configuration . The detailed circuitry employed in this integrated circuit and performance data are readily available from the manufacturer. [11] The price of this device is not considered to be excessive at the present time.



Note: Letter-number combination inside square indicates circuit board by the letter and the terminal of the board by the number.

Figure 6
Preamplifier Schematic

7. MIXER CIRCUIT

The transfer characteristic of field-effect transistor (FET), made by the diffusion process, has a square-law relationship between the drain to source current, I_{ds} and the gate to source voltage, V_{gs}. It is expressed as [25]

$$I_{ds} = I_{DSS} \left(1 - V_{gs}/ V_P \right)^2 \tag{1}$$

where I_{DSS} is the saturation drain current when the gate is shorted to the source ($V_{GS} = 0$) and V_P is the pinch off voltage. For mixer operation let V_{gs} be represented as the sum of two sinusoidal voltages both of which can be simultaneously impressed on the gate of an FET or one impressed on the gate and the other on the source of the FET. Either process will cause mixing operation and V_{gs} as defined below holds true for both cases

$$V_{gs} = V_{GS} + V_s \cos w_s t + V_o \cos w_o t \tag{2}$$

where V_{GS} is the bias gate to source voltage. $V_s \cos w_s t$ represents the source or voice sound while $V_o \cos w_o t$ is the sinusoid generated by the local oscillator which is applied to the gate or source of the FET. Substituting (2) into (1) and expanding, we obtain

$$\begin{aligned} I_{ds} = \frac{I_{DSS}}{V_p^2} \Big\{ & V_P^2 + V_{GS}^2 + \tfrac{1}{2}V_s^2 + \tfrac{1}{2}V_o^2 \\ & -2(V_P - V_{GS}) (V_s \cos w_s t + V_o \cos w_o t) \\ & + \tfrac{1}{2} V_s^2 \cos 2w_s t + \tfrac{1}{2} V_o^2 \cos 2w_o t \\ & + V_s V_o \left[\cos(w_s + w_o)t - \cos(w_s - w_o)t\right] \Big\} \end{aligned} \tag{3}$$

The drain current has DC components plus six individual frequencies as a result of the square law mixing of an FET. This response shows that only frequencies of the form w_s, w_o, $2w_s$, $2w_o$, $w_s + w_o$, and $w_s - w_o$ are obtained while other frequencies of the form $mw_s \pm nw_o$, which must be suppressed in conventional mixers, are greatly reduced with an FET mixer circuit. [2]

The frequency component of the drain current which is of interest is $V_sV_o\cos(w_s - w_o)t$. It is separated from the other components by coupling to the mixer output a DC blocking capacitor followed by a low-pass filter which has a cut off frequency w_{fil} such that $w_s - w_o < w_{fil} <$ both w_s and w_o.

Initially, a dual gate metal oxide semiconductor (MOS) FET was selected as being particularly well suited for use in mixing two audio frequencies. Eight 3N141 MOS-FET's were ordered but due to excessive delay in receipt of these devices, it was necessary to design and build a mixer using a single gate FET already in stock in the school electronics issue room. This device requires that the voice signal be impressed on the gate while the local oscillator signal is applied to the source terminal. Several types of FET's available from the issue room were tested for mixing action in the circuit shown in figure 7a. The 2N3819 proved to be the most satisfactory device. Its transconductance as a function of gate to source voltage is quite linear over the range from zero V_{GS} to pinch off voltage V_P. This characteristic enhances the mixing action of an FET. [20]

The local oscillator used in AESTR is a URM-127 signal generator. It has an output impedance of approximately 100 ohms and can deliver a signal ranging from the microvolt range to a maximum of 10 volts.

In designing the mixer circuit the author relied on the manufacturer's data sheet for the 2N3819 FET. It is an N-channel device with $V_P = -8$ volts, $I_{DSS} = 10$ ma and an average

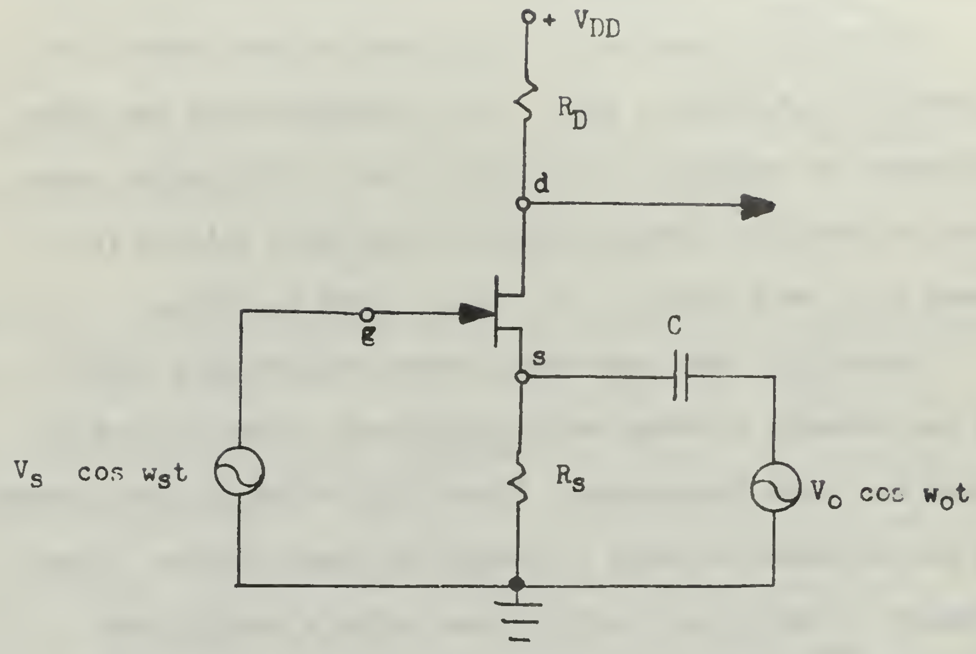


Figure 7a

FET Mixer Network

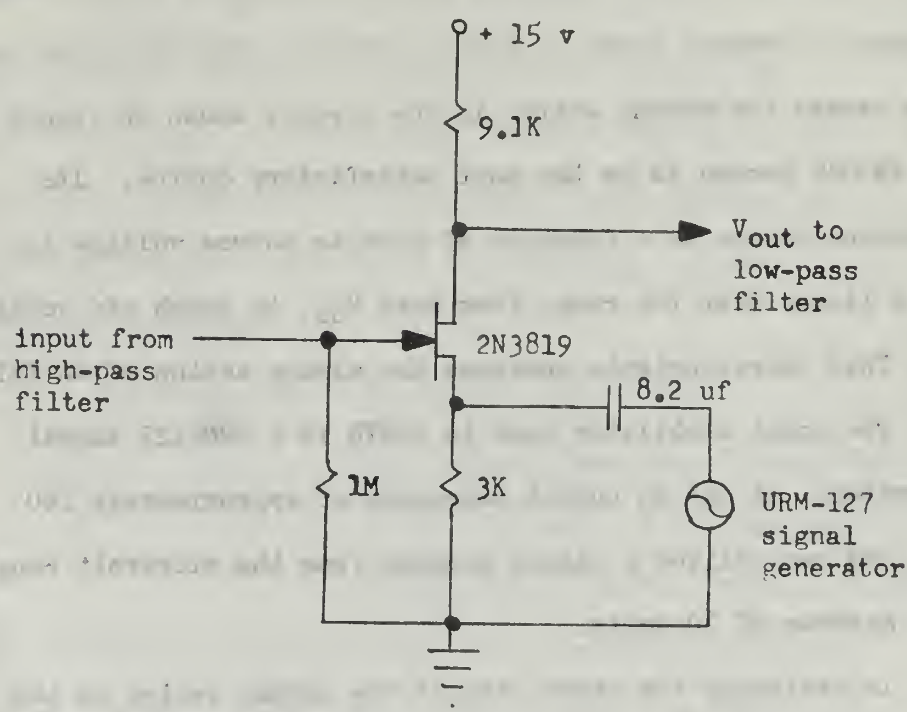


Figure 7b

AESTR Mixer Schematic

transconductance of 4000 micromhos for zero gate bias. The DC drain current I_D was selected to be 1 ma. and the mixer circuit was designed to give a voltage gain of 10. These conditions were incorporated into the circuit design [25] and values were obtained such that R_S = 5.5 kohms and R_D = 8.1 kohms. The network in Figure 7a was constructed and the components subsequently modified to the circuit of Figure 7b to obtain optimum mixing.

Successful mixing of any two audio frequencies is accomplished by means of this circuit with no lower limit on the input and local oscillator voltages. An upper limit of 1.5 volts peak to peak for the signal and local oscillator voltages cannot be exceeded; otherwise the output is clipped. Optimum operation of this mixer circuit is set for an input of approximately 0.8 volts peak to peak. Above this voltage, the follow-on filter circuits begin to give spurious outputs due to sweeping of either the voice oscillator or local oscillator across the frequency spectrum. This effect is noticeable on the F1 and F2 voltmeter indicators and masks the frequency response of the filters.

The 2N3819 FET's have consistently performed the mixing operation on a daily basis during the entire period covering the design and testing of the formant indicators. These particular FET's are highly recommended both for their reliability and usefulness in audio mixing circuits.

As an epilog to the mixer design realization, the 3N141 MOS-FET's did arrive finally. Other students have had limited success in using these devices for mixing. Special care must be exercised in using them, especially with regard to preventing any external high voltages (static charges, etc.) from accidentally damaging the devices.

8. FILTER CONSIDERATIONS

Many types of network designs will yield either low-pass, high-pass or band-pass frequency filters. The networks may be synthesized using only passive elements [18, 19] or in addition to resistive or capacitive components, incorporate a radio tube, [35] transistor, [15] or an integrated circuit operational amplifier. [4,5,14] When a filter design calls for cutoff frequencies below 100 Hz, several considerations tend to indicate that an active RC filter circuit is the most desirable type. Table 5 lists the relative characteristics of passive and active filters with cutoff frequencies below 100 Hz.

Active network synthesis can be classified in a number of ways, depending on the purpose of active elements and the network configuration. The three main types of active synthesis consist of a. <u>Classical Amplifier Design</u> where the active element is part of the parameters of the network. b. <u>Feedback Systems</u> where feedback theories are used to synthesize poles and zeros of a network function. In this case active elements are used as isolation or amplification devices, or as functions of operational amplifiers. c. <u>Modification of Passive Synthesis</u> where techniques of passive synthesis are used to realize portions of a network that are connected together by active elements. In all three categories listed, the active elements are used mainly as controlled-source devices which perform functions of subtraction , negative-constant multiplier or inversion. They can be treated as black boxes performing their prescribed mathematical functions. [38]

The ideal low-pass filter with unity transmission below

TABLE 5

COMPARISON OF PASSIVE AND ACTIVE FILTER CHARACTERISTICS
FOR LOW FREQUENCY APPLICATIONS

Passive RLC or RC Filters	Active RC Filters
1. Inductors tend to be expensive, harge, heavy and susceptible to hum (A-C line frequency) pickup.	1. An inductor can be replaced by an active circuit which has an appropriate input impedance.
2. For filters consisting of only resistors and capacitors, the poles and zeros of the driving-point immittance functions of RC filters are restricted to the negative, real axis of the s plane, and the same is true for the poles of the transfer functions.	2. It is possible to use an active element as a cathode or emitter follower or an operational amplifier in synthesis of the filter.
3. A maximum attenuation of 6 db per octave can be obtained with each individual RC filter section.	3. It is possible to realize driving-point functions and transfer functions with no restriction on the poles and zeros.
4. RC filters exhibit attenuation of the signal in the designed pass band.	4. Positive pass band gain can be designed into the circuit.
	5. Simpler network configurations at lower cost can be achieved.

and zero above a certain frequency, with no phase shift in the pass band, is unattainable in the real world. Three approximations to the ideal filter can be realized by means of the Butterworth, Bessel or Chebyshev filters. [14]

The Bessel filter exhibits maximally flat time delay (linear phase) and therefore sometimes is used as a time delay network. Its amplitude response in the pass band is monotonically decreasing rather than flat. Its rate of fall beyond cutoff is less than the Butterworth or Chebyshev filters.

The Chebyshev class of filters have an equal magnitude ripple in the pass band and maximum rate of fall beyond 3 db cutoff. The response of the filter at the cutoff frequency is always that of a minimum of the ripple. The allowable degree of ripple in the pass band can be accounted for in the filter design.

The Butterworth filter is obtained by locating the poles of the network in accordance with the zeros of the Butterworth Polynomial. The normalized transfer function is of the form

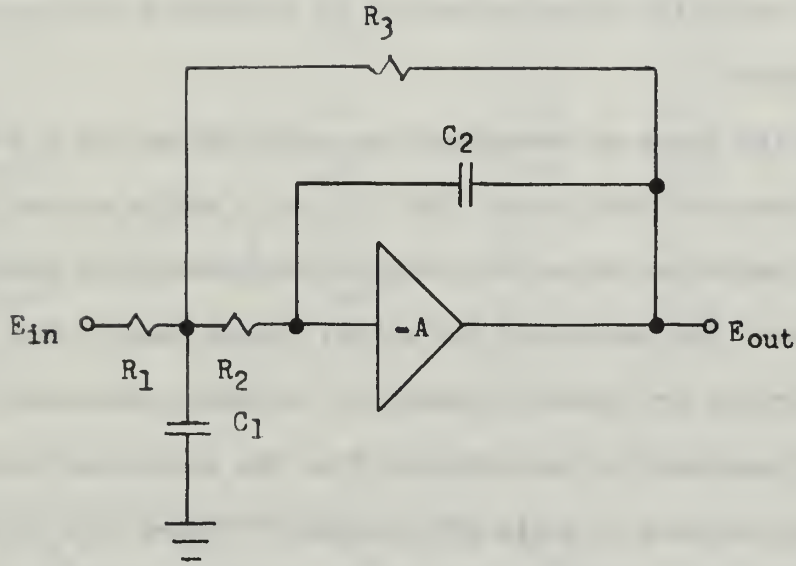
$$/Z_{12}(jw)/^2 = \frac{1}{1 + w^{2n}}$$

where n is the number of poles in the network and w is the ratio of frequency of interest to cutoff frequency. The filter has a maximally flat amplitude response in the pass band and the slope of rolloff outside the pass band increases directly with the number of poles in the transfer function. The response falls off at approximately a constant 6n db/octave. The phase characterisitics of the Butterworth filter are not very linear. The time delay varies as a function of frequency.

9. LOW-PASS FILTER DESIGN

The F1 low-pass filter and the F2 low-pass filter in AESTR are identical circuits. Each filter is a four pole Butterworth response circuit with discreate cutoff frequencies of 10, 15, 30 and 60 Hz. The Rauch type filter network is selected since the circuit values are rapidly calculated for multiple filter sections by using the normalized tables contained in Foster's paper. [14] Also this network can be modified to provide a continuous variable cutoff frequency or have positive gain by modifying the resistive elements of the circuit. [28] In AESTR, the filters have unity gain and the cutoff frequencies are established by switching various capacitor values into the network while maintaining all resistor values at a constant value of 10K. The author decided to vary the capacitors rather than the resistors to control cutoff frequencies because of hardware considerations. As more data and experience is gained in the operation of AESTR, it may be desirable to design positive gain and continuous variable cutoff frequency into the filters based on recommendations of the speech therapists. Each filter is mounted on a separate circuit board and modifications can be accomplished without changing the internal chasis wiring.

The Rauch filter basic building block is a single section which has two poles in the complex frequency plane. Its schematic and transfer function are shown in figure 8. Two cascaded sections are required to obtain a roll-off of 24 db per octave for frequencies above the cutoff frequency. A 25 uf coupling capacitor is inserted between sections to block D.C. components while a 10K shunt resistor is inserted at the input of each



$$\frac{E_{out}}{E_{in}} = \frac{-\dfrac{1}{R_1R_2C_1C_2}}{s^2 + \dfrac{R_2R_3 + R_1R_3 + R_1R_2}{R_1R_2R_3C_1}\, s + \dfrac{1}{R_2R_3C_1C_2}}$$

Figure 8

Two Pole Rauch Low-Pass Filter and Transfer Function section to provide a D.C. return path to the base of the inverting input transistor enclosed in the Fairchild uA 709 operational amplifier. This resistor also develops the required input voltage necessary for proper filter response. All filter network resistors are fixed at 10 K ohms to provide an adequate filter impedance match to the mixer output and to determine practical capacitor values which can be obtained for fabrication of the network.

From the table of normalized capacitor values for a Butterworth filter with four poles, [14] , it is a simple matter to calculate capacitor values for various low-pass filter cutoff frequencies. The calculated and actual values used in the AESTR apparatus are listed in table 6. Although the actual capacitor component values deviated from the calculated values, the filter response is quite satisfactory. Figure 9 is a plot of the frequency response curves of the low-pass filters in AESTR.

The various capacitors are mounted on a five pole two gang switch which is operated from the front panel of AESTR. The ten inch cable wires between capacitors and circuit boards do not contribute any noticable adverse effect on the filter response.

A zero output response is observed for zero beat frequency output of the mixer stage due to the coupling capacitors of the filter. This effect does not affect the purpose for which AESTR is to be used since it is practically impossible for a person to hold his vowel formants exactly on frequency with the local oscillators. The continuous deviations of the formants

TABLE 6

CAPACITOR VALUES OF FOUR POLE BUTTERWORTH RESPONSE RAUCH TYPE LOW-PASS FILTER WITH ALL RESISTOR VALUES SET AT 10K OHMS

Capacitor Value (uf)		Cutoff Frequency (Hz)			
		10	15	30	60
C_1	computed	6.25	4.17	2.08	1.04
	actual	8.2	4.0	2.0	1.0
C_2	computed	0.41	0.27	0.14	0.068
	actual	0.4	0.22	0.13	0.068
C_3	computed	2.58	1.73	0.86	0.43
	actual	4.0	1.5	0.8	0.4
C_4	computed	0.98	0.65	0.33	0.164
	actual	1.0	0.8	0.4	0.168

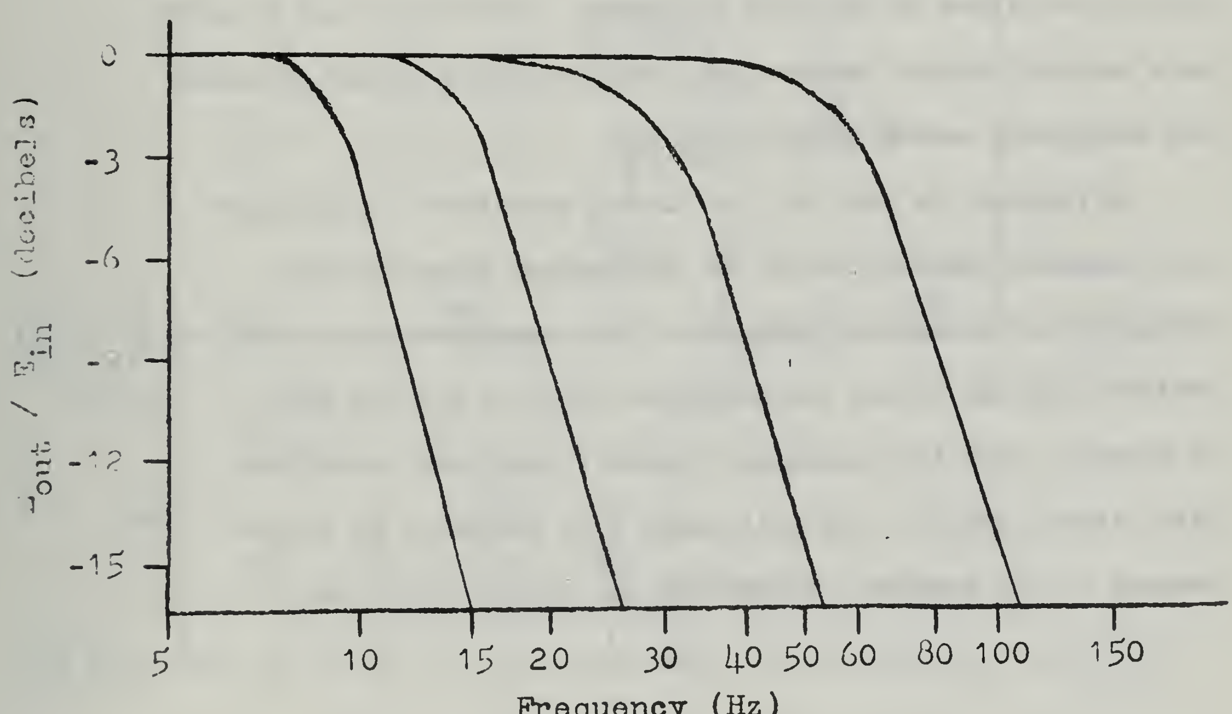


Figure 9 Low Pass Filter Response Curves

are sufficient to cause beat frequencies which will be present in the pass bands of the filters.

The filter output is passed into an amplifier using a 1 K ohm input resistor and a 1 Megohm potentiometer in the feedback circuit across a uA 709 operational amplifier. The potentiometer control is designated as "filter sensitivity" on the front panel of AESTR.

As stated previously, two identical low-pass filters are contained in the AESTR system. One filter responds to the first formant beat frequency and the other responds to the second formant beat frequency generated in their respective mixer circuits. Figure 10a is a schematic of the complete filter network while Figure 10b is a schematic of the beat frequency amplifier which drive a 0-10 volt rectifying voltmeter. Several typed of meters were considered for use as visual indicators of the beat frequency. The meters used in AESTR were selected simply because they were available in the stockroom and adequately served AESTR's purpose.

In Figures 10a and 10b, the uA 709 operational amplifiers are frequency compensated in the same manner shown in the preamplifier schematic of Figure 6. The components have been omitted from the filter and amplifier circuits for the sake of clarity. Also the schematics identify terminals associated with circuit board B. Circuit board C is identical to B with respect to all terminal connections and component values.

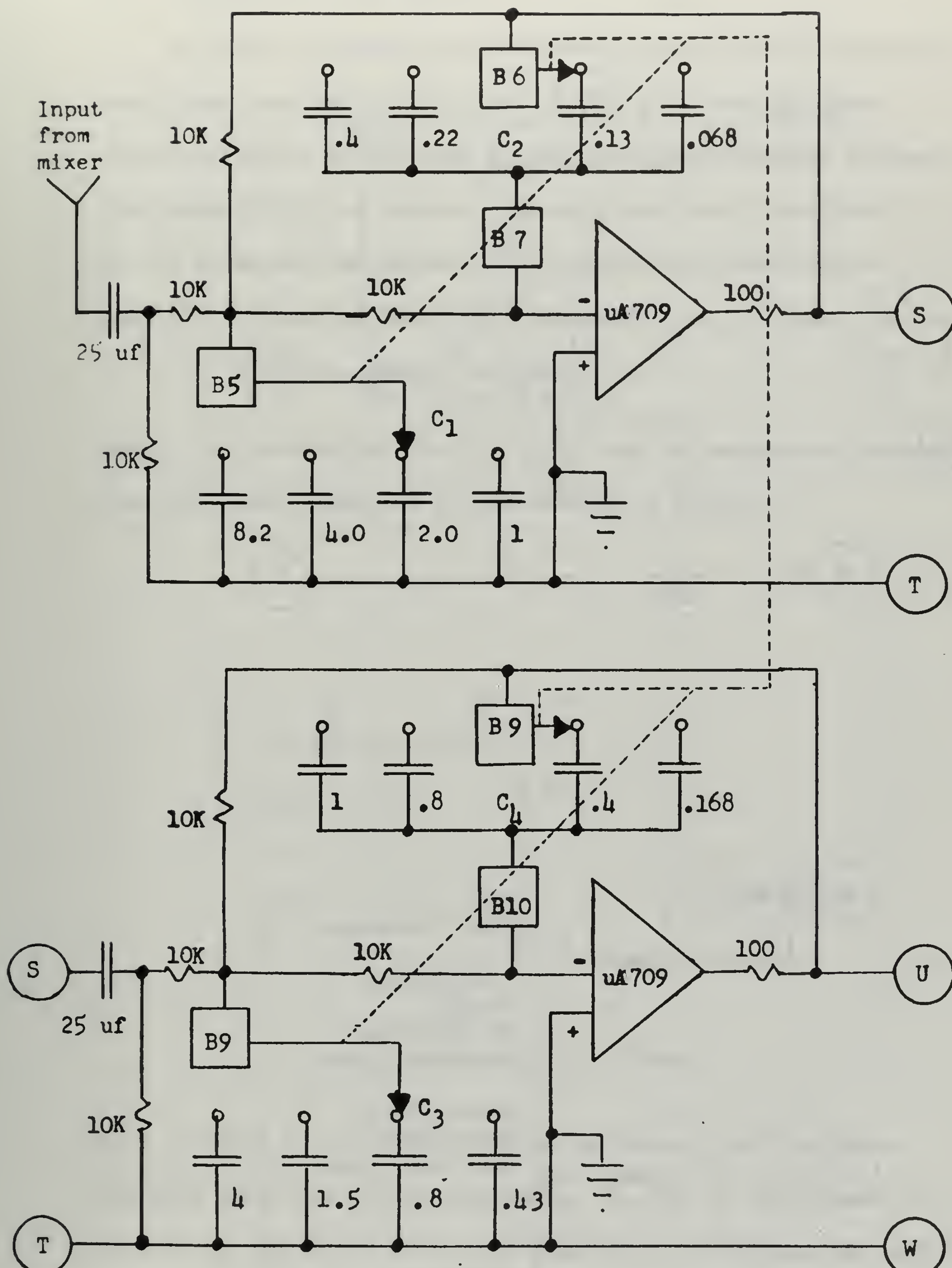


Figure 10a. Four Pole Rauch Low-Pass Filter Schematic

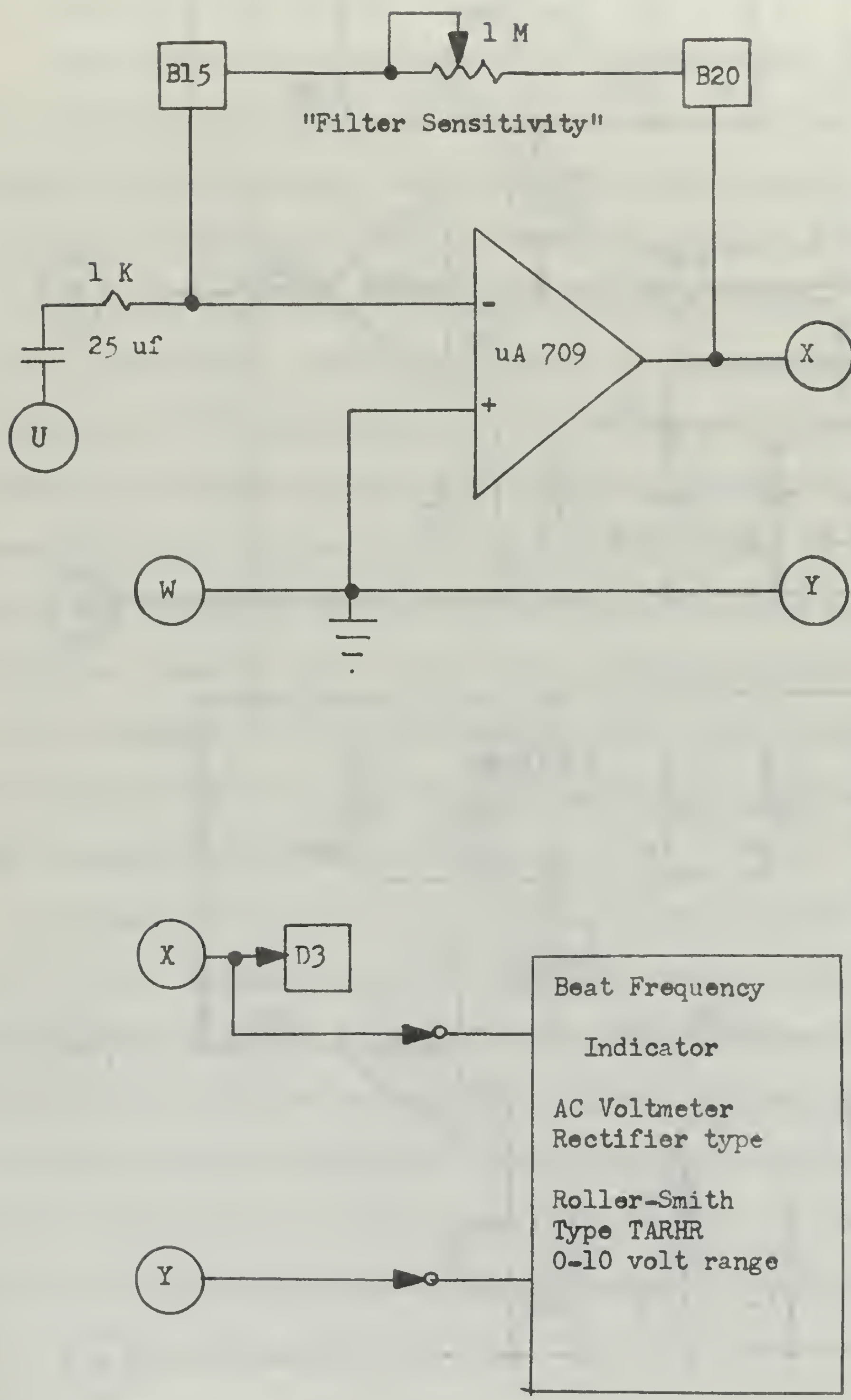


Figure 10b

Low-pass Filter Gain Schematic and Beat Frequency Indicator

10. HIGH-PASS FILTER DESIGN

In order to prevent the fundamental frequency of the vocal bands from passing directly through the mixer and low-pass filter circuits, a high-pass filter network is inserted between the preamplifier and mixers. Its configuration is realized by the Salen and Key method. [35] A high-pass filter has a normalized frequency transfer function of

$$\frac{E_{out}}{E_{in}} = \frac{S^2}{S^2 + ds + 1} \qquad (4)$$

where d is the damping factor. This type of response is obtained from the basic high-pass filter network of Figure 11.

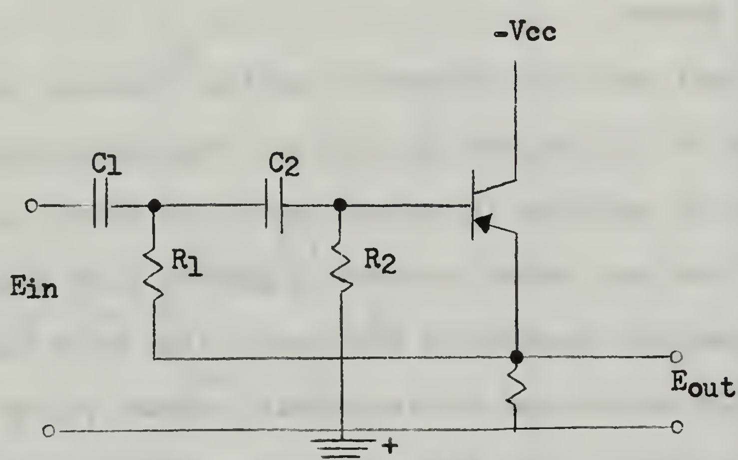


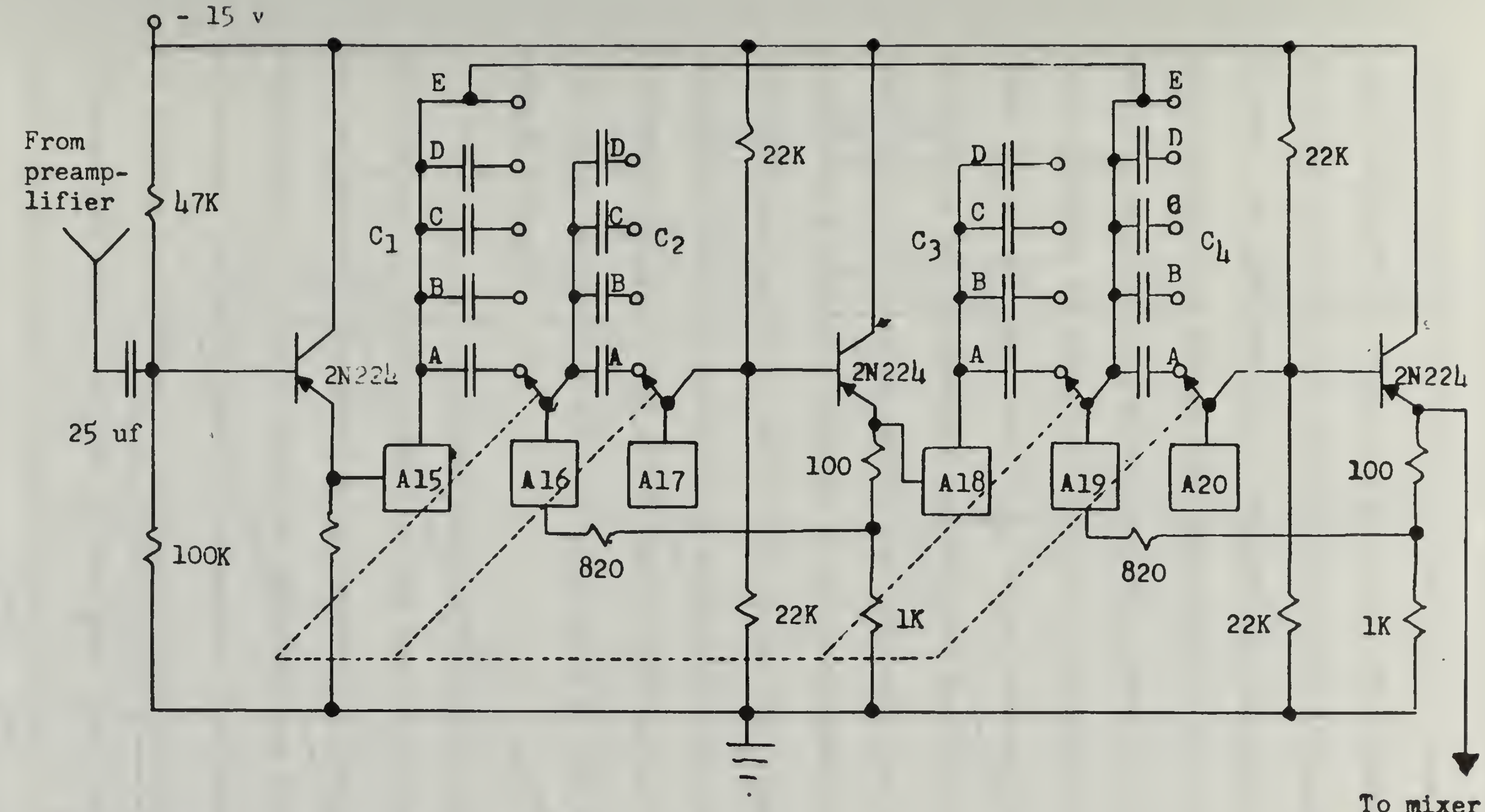
Figure 11.

Basic High-Pass Filter Network

Such a filter will give a 12 db per octave roll-off for frequencies below the cutoff frequency. R_1, R_2, C_1 and C_2 and the gain of the emitter follower act together to determine the cutoff frequency and the shape of the response curve during the transition from the stop band to the pass band. In the actual circuit, R_2 is equal to the resistance of three parallel resistors. These are the two bias resistors and the input resistance of the 2N226 transistor.

The AESTR high-pass filter schematic is shown in Figure 12. An emitter follower drives the high-pass filter stage which consists of two cascaded sections to yield an expected attenuation of 24 db per octave. [16] The individual sections do yield a Butterworth response of 12 db per octave roll-off, but, when cascaded together, a total roll-off of only 20 db per octave is realized with an additional +3 db hump at the corner frequency. The actual response shown in Figure 13 is considered satisfactory for the pitch elimination function in AESTR's system.

Note that the pitch eliminator has four discrete cutoff frequencies of 75, 190, 450 and 1050 Hz. The desired cutoff is obtained by switching in various capacitors mounted on a five pole, two gang switch attached to AESTR's front panel. The fifth position permits the high-pass filter to be bypassed so that AESTR can be used to discriminate between voiced and unvoiced consonants. This feature was incorporated into the apparatus after Dr. Gray operated a breadboard version of the system and suggested that a "pitch" or "no pitch" capability be incorporated into AESTR.



Note: a. $C_1 = C_2 = C_3 = C_4$

b. Capacitor values for switch positions are

Position	A	B	C	D	E
Capacitor (uf)	.5	.4	.1	.047	Bypass

Figure 12. Two Section High-Pass Filter Schematic

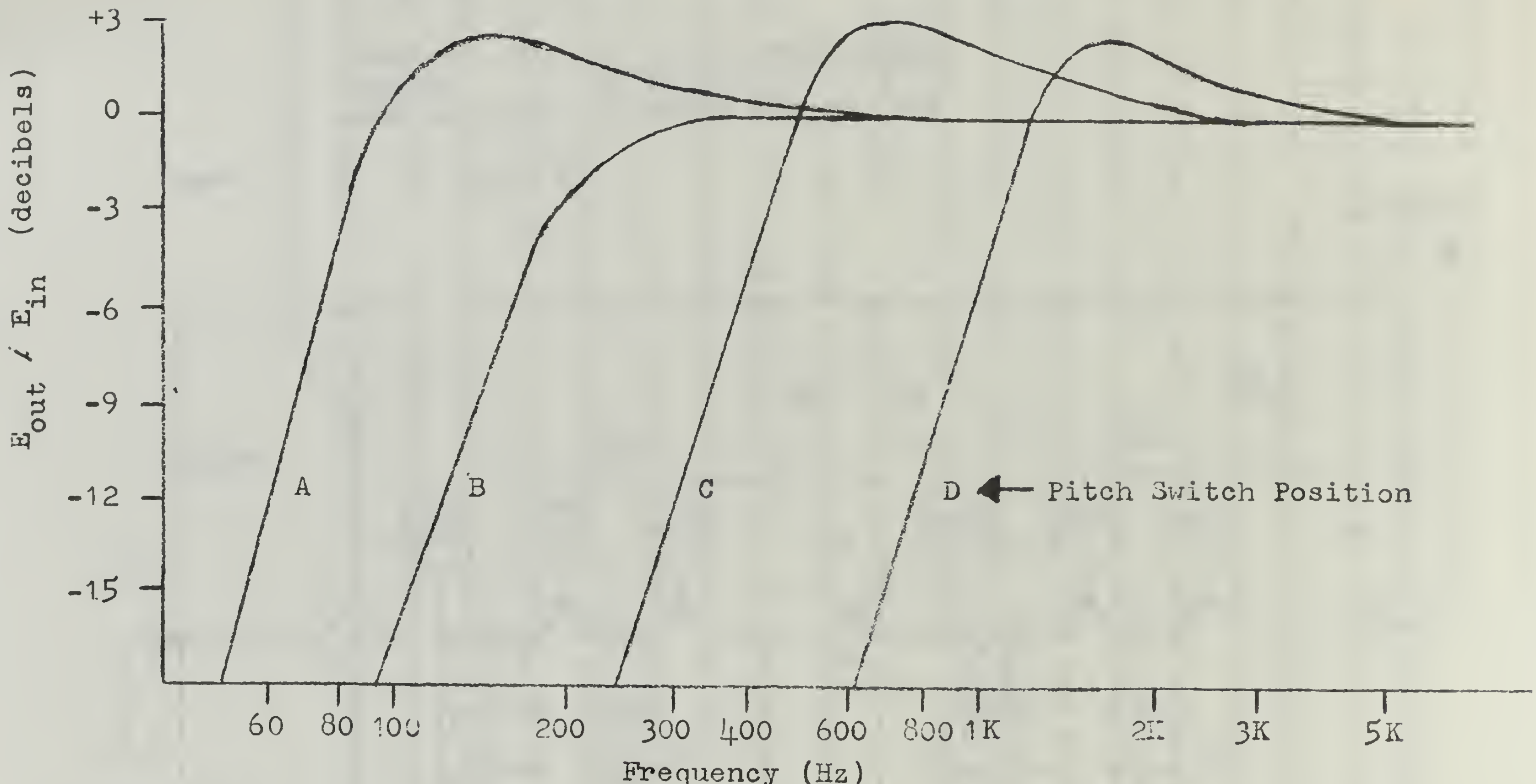


Figure 13. High-Pass Filter Response for Pitch Selector Switch Positions

11. DECISION AND RESPONSE CIRCUIT DESIGN

The beat frequency output of the first and second formant filters is applied to terminals 3 and 20 of circuit board D whose schematic is whown in Figure 14. These waveforms are Half-wave rectified and smoothed by a low-pass passive RC filter. The resultant D.C. voltages are impressed on the two input gates of an AND circuit. When both F1 and F2 beat frequency rectified voltages are simultaneously present and also of sufficient magnitude to cause +7.5 volts D.C. to appear on each diode of the AND gate, the diodes become reverse biased thereby directing a 400 microampere current into the base of the 2N2924 transistor. This action drives the transistor into saturation, permitting a collector current of 30 milliamperes to flow through the relay coil, which acts as the load for the circuit, and closes the relay contacts. A zener diode is inserted at the base terminal of the transistor to prevent the transistor from being switched on when only one diode of the AND gate is reverse biased.

The relay is a stockroom surplus item which operates on 14 volts and 25 milliamperes. It has two sets of contacts. One set activates the green panel "correct" light and the other set connects a 115 volt supply to the appliance socket mounted on the rear chasis of AESTR. The Monterey Institute for Speech and Hearing does have a 115 volt relay operated device which dispenses M&M candy disks to children when they perform desired tasks. AESTR is able to operate this dispenser or any other 115 volt device in response to the desired articulation of the child.

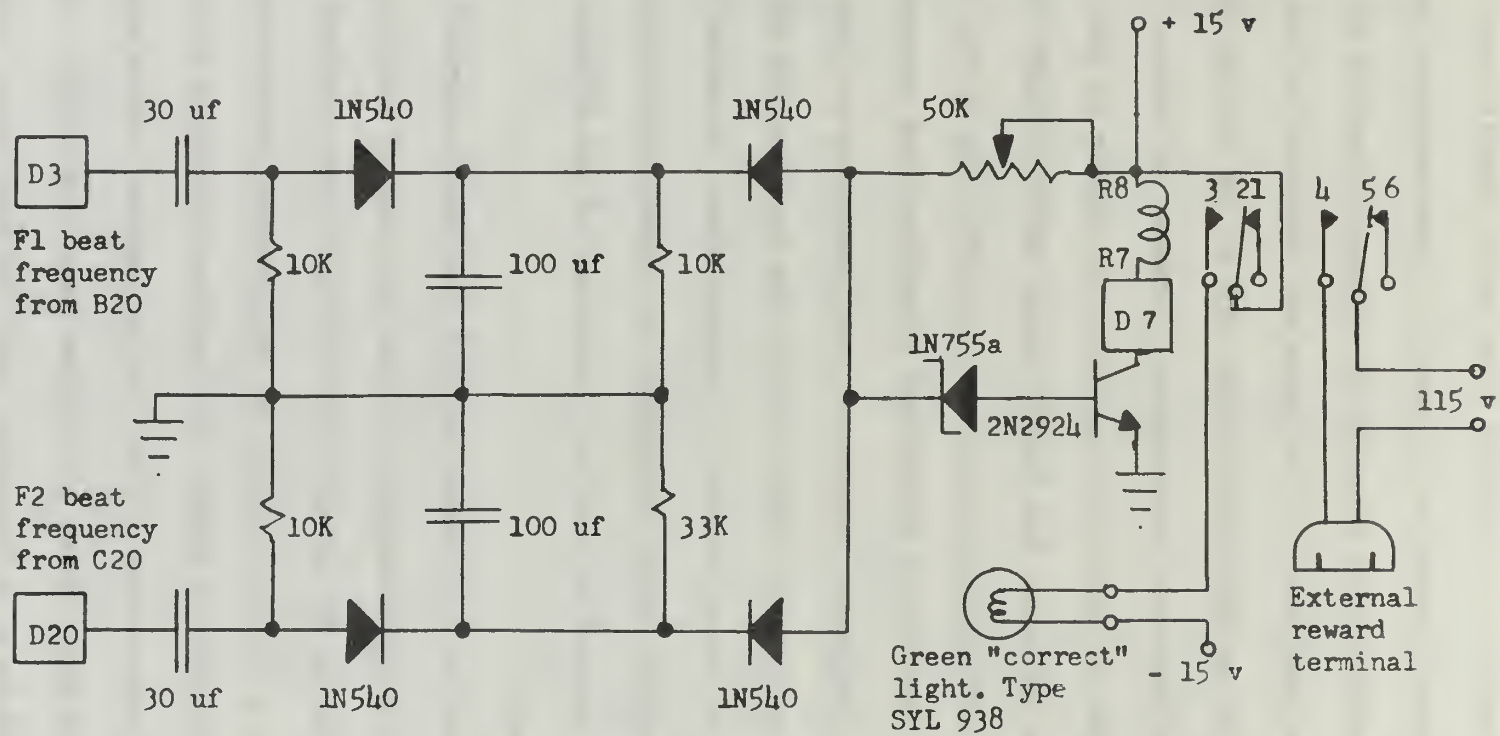


Figure 14. Decision and Response Circuit

12. FABRICATION

Economy and availability of supplies dictated construction of AESTR. All components are housed in an aluminum case 16" wide, 12" deep and 10" high. The control panel is inclined 20° from the vertical so that the values of the control settings can be read with greater ease. The case was handmade in the student metal shop. In addition the control panel was rubbed with emery paper until the metal acquired a satin finish.

The chasis for circuit components has four 22 terminal sockets which accept the standard $4\frac{1}{2}$" by 6" circuit boards. Also mounted on the chasis is an 11 pin socket for the power supply package, mounting holes for the relay plus an octal socket for power distribution cables and a 27 pin socket for signal distribution cables which originate from the components mounted on the rear of the control panel. Fusing is provided for circuit protection.

The circuit boards are identified by letters which are:

Board A Preamplifier, High-Pass filter, Mixers

Board B F1 Low-Pass filter, amplifier

Board C F2 Low-Pass filter, amplifier

Board D Rectifier, AND circuit, transistor switch

The functional segregation of the circuit boards permits future changes to the circuitry by simply replacing an entire board. It should not be necessary to change the internal wiring of the chasis for such modifications.

Original circuit boards used for mounting of components were the etched contact plugboards Vector #838PWE. They are considered to be restrictive in flexibility. The Vector

#838 semi-etched boards proved to be more versatile. Components are mounted easily and securely with the aid of metal washers riveted on the holes through which the lead wires pass through to the other side of the board. Additional holes must be drilled into the board to accomodate the integrated circuit octal socket. Learning how to properly mount components so as to conserve space, minimize leads and avoid ground loops is considered by the author to be a very useful and important aspect of this thesis.

The electronic circuits of AESTR require 30 milliamperes on both the plus and minus 15 volt supply terminals. When the relay and "correct" light are activated, the current drain increases to 95 milliamperes on both supply terminals. The power is supplied by a Power Mate Power Supply, Model DRA16-.2/16-.2. Its regulated output can be set between 15 and 17 volts and is rated to provide 200 milliamperes on the plus and minus terminals. The voltage regulation is excellent even during sudden current level changes when the light and relay activate. Figure 15 shows the power distribution in AESTR.

As stated previously, the filter capacitors are mounted on five pole, two gang switches. These components are located longitudinally around the periphery of the switches so as to economize on space and also obtain structural support.

Trouble shooting the system after AESTR was completely wired consumed many hours. A component value error and cable error required correcting before successful operation of the assembled machine could be achieved.

Numerous minor problems were encountered in the fabrication

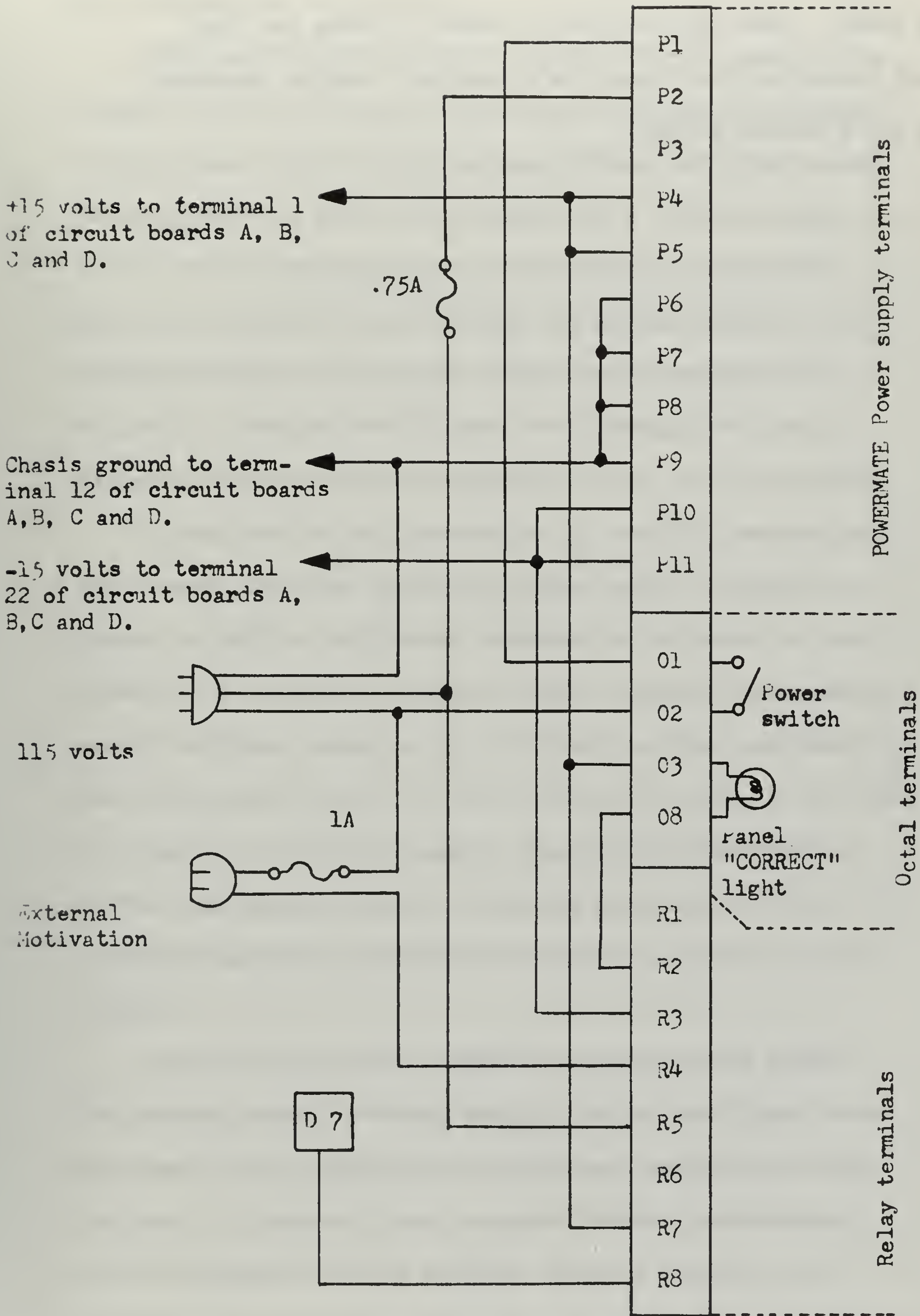


Figure 15.

AESTR Power Distribution

of AESTR. These difficulties did serve to prove the fact that transition from theory to a practical working apparatus is not a trivial matter.

13. PRELIMINARY TEST RESULTS

AESTR was initially tested during the final phase of its design stage by Dr. Gray at the school electronics laboratory. At that time, the F1 and F2 low-pass filters had fixed cutoff frequencies of 50 and 100 Hz respectively. His evaluation of the machine indicated that the pass band of the filters had to be reduced in order to have the machine properly discriminate between the closely related voiced sounds such as ER and E. Therefore the filters were redesigned to have a series of discrete cutoff frequencies of 10, 15, 30 and 60 Hz.

During this initial evaluation, it was also learned that air streams impinging on the microphone cause a transient response in AESTR of sufficient magnitude to activate the relay circuit. To avoid such a type of false response, the speaker should hold the microphone in a vertical position approximately four inches away from and slightly below his lips. In the case of a child, a microphone headset type configuration similar to the kind commonly worn by telephone operators would keep the microphone properly positioned relative to the mouth of the speaker.

After its fabrication, AESTR was tested by the author. The machine control settings obtained for an adult male voice and female voice articulation of the vowel sounds are listed in Table 7. These settings represent the best values which could be obtained for the smallest spectral window in the F1-F2 plane. In all cases, the first formant of the vowel was readily located with minimum sweeping of the F1 local oscillator. The second formant was more difficult to locate for

TABLE 7

AESTR CONTROL PANEL SETTINGS FOR VOWEL SOUNDS SPOKEN BY AN ADULT MALE VOICE AND AN ADULT FEMALE VOICE

Speaker	Vowel	Volume control	Pitch control	F1 LPF control	F1 Sen-sitivity	F2 LPF control	F2 Sen-sitivity	F1 Loc. Osc. (Hz)	F2 Loc. Osc. (Hz)
TK male	IY	10	B	60	10	60	10	370	2200
	I	10	B	60	10	60	10	390	1740
	E	7	B	30	5	60	5	520	1600
	AE	7	B	30	5	30	7	680	1600
	A	7	B	60	5	60	3	680	1000
	OW	7	B	30	3	60	3	500	860
	U	7	B	30	2	60	4	450	760
	OO	7	B	30	4	60	3	300	800
	UH	7	B	30	4	60	4.5	520	1090
	ER	7	B	60	2.5	60	4.5	420	1200
SK female	IY	6	A	60	2	60	10	420	2700
	I	5	A	30	5.5	60	10	450	2250
	E	5.5	A	30	5.5	60	10	490	2420
	AE	5	A	30	6	60	3.5	670	1700
	OW	2.5	A	30	6.5	60	3	750	920
	U	2.5	A	30	6.5	60	5.5	460	1000
	OO	6.1	A	30	3	15	9	240	715
	UH	9.5	C	30	8	15	9	500	1450
	ER	8	B	30	8	30	9	520	1500

vowel sounds IY, I, and ER. The F2 local oscillator must be swept across its frequency range three or four times before the operator is certain that the F2 frequency has been located. This is to be expected since the amplitude of the second formants is lower than the first formant for all vowel sounds.

Pitch measurements were made according to the procedures stated in section 5. Pitch frequencies are rapidly determined and do show a variation with the vowel sounds as indicated in Table 2.

TABLE 8

AESTR PITCH MEASUREMENTS OF AN ADULT MALE VOICE FOR VOWEL SOUNDS

Vowel	IY	I	E	AE	A	OW	U	OO	UH	ER
Pitch (Hz)	110	117	98	98	96	90	112	108	104	100

AESTR is now on loan to the Monterey Institute for Speech and Hearing for field testing. Their preliminary operation of the apparatus in conjunction with an M&M candy dispenser revealed a new problem. Candy disks were being dispensed at a very rapid rate since the relay opened and closed every time the voice quived in and out of the desired sound spectral window. Therefore, to make AESTR provide only one reward item with each sustained sound, the AND circuit was modified to have a 250 millisecond delay before closing the relay contacts, and once closed, the relay would not open for two seconds. This modification consisted of choosing the correct shunt caapacitor values in the half-wave rectifier portion of the decision and response circuit. A nominal value of 100 microfarads working with the resistive elements of the circuit develops build-up and decay time constants to meet the operating specification for the relay.

Dr. Gray and his associates tested AESTR for its ability to discriminate the individual vowel sounds. The preliminary results indicate that the machine, for certain vowels, will give a positive response to not only the targeted vowel but also to certain other vowel sounds. For example, AESTR can be set to respond to OW and it will perform properly such that the speaker is unable to cause a positive machine response with any vowel sound other than OW. However, if AESTR is targeted for the central vowel sound ER, the machine will respond to ER plus the phonemes A, OW, U, OO and UH. The apparent cause for this undesirable multi-sound response is due to the fact that ER has a relatively low intensity level for its first and second formants when compared to back vowels, especially OW. Unfortunately, the therapist has a greater need to teach the ER rather than OW to speech handicapped children. To improve AESTR's ability to respond strictly to the ER sound, Dr. Gray and the author varied the "pitch" control settings. The attempt indicated that some improvement could be made if the "pitch" control is set to position "D". Now the machine will respond only to ER and OW. The OW vowel continues to mask all other vowels since it does contain the greatest amount of energy throughout the audio frequency spectrum.

A different approach was tried to overcome the ER ambiguity response of AESTR. Both the F1 and F2 local oscillators were set to the second formant frequency of 1480 Hz while the "pitch" control remained in position "D". The volume control was set to a value of 3 and the sensitivity controls were set to a value of 4. In this state, the machine would respond only to the ER sound for a majority of trials. This can be explained by noting that OW has both its F1 and F2 frequencies below 1 KHz which are attenuated by the high-pass filter and the harmonic components of OW near 1480 Hz are insufficient to cause a positive response of the machine. Now, when a speaker makes the ER sound, its second formant (near 1480 Hz) is not attenuated by the high-pass filter and will provide a strong beat frequency out of both F1 and F2 filters thus causing AESTR to give a positive response. This type of machine operating procedure will be investigated further and extended to take advantage of the third formant information associated with each vowel.

A speaker is able to cause AESTR to give a positive response when he greatly increases the intensity of his voice. The author recommends that some type of distortionless speech compressor be inserted between the microphone and preamplifier. Commercial devices are readily availabe to control the microphone peak loudness yield.

A human limitation prevents AESTR from being operated for more than 15 minutes by one speaker. After a person has been producing voiced sounds for this period of time, he will start becoming hyperventilated and experience dizziness. The effect is analogous to a person blowing up a large balloon. Dr. Gray is giving consideration to this factor and will develop a clinical testing procedure to avoid hyperventilation of the speaker.

14. CONCLUSIONS

The prototype apparatus does perform electrically in the manner it was designed to operate but this does not imply that AESTR is performing in a totally satisfactory manner from the viewpoint of the speech therapist. AESTR is considered to be approximately 50% successful in meeting the needs of the therapist. With more operating data obtained from the machine in future months, it is hoped that additional design criteria can be established to improve AESTR's performance.

In addition to aiding speech handicapped children, AESTR has potential applications to aid persons trying to learn foreign vowel sounds. Also this apparatus can be used in an auxiliary manner to measure tones of musical instruments such as pianos or organs with a high degree of accuracy.

Speech processing and especially specific analysis of spectral components of voiced sounds is a challenging task from an engineering viewpoint. This fact became very apparent from what appeared to be a very straight forward thesis subject.

15. BIBLIOGRAPHY

1. Bach, R. E. "Selecting R-C Values for Active Filters" Electronics, v. 33, 13 May 1960, pp. 82-85.

2. Bosenber, W.A. "Transfer Characteristics of Field-Effect Transistors". RCA Review , v. 23, December 1963,pp.688-699.

3. Buchanan, C.D. A Programmed Introduction to Linguistics. Boston, Mass.: D.C. Heath & Co., 1965.

4. Burr-Brown Research Corporation. Handbook of Operational Amplifier Applications. Tucson, Arizona, 1963.

5. Burr-Brown Research Corporation. Handbook of Operational Amplifier Active RC Networks. Tucson, Arizona, 1966.

6. Denes, P.B. and E.N. Pinson. The Speech Chain. Baltimore, Maryland: Waverly Press, 1963.

7. Donaldson, R.W. "Communication via the Auditory, Tactile and Kinesthetic Senses." IEEE Transactions on Information Theory, v. IT-13, January, 1967, pp. 11-21.

8. Dudley, H. "The Carrier Nature of Speech." The Bell System Journal, v. 19, October, 1940, pp. 495-515.

9. Fairbanks, G. Experimental Phonetics. Urbana, Illinois: University of Illinois Press, 1966.

10. Fairbanks, G. Voice and Articulation Drillbook. Second Edition, New York: Harper & Row, 1960.

11. Fairchild Semiconductor. Linear Integrated Circuits Applications Handbook. Written and edited by J. Giles, Mountain View, California, 1967.

12. Ford, R.L. and F.E. Girling. "Active Filters and Oscillators using Simulated Inductance." Electronics Letters. v. 1 , December, 1965, p.287.

13. Forgie, J.W. and C.D. Forgie. "Results Obtained from a Vowel Recognition Computer Program". The Journal of the Acoustical Society of America. v. 31, November, 1959, pp. 1480-1489.

14. Foster, E.J. "Active Low-Pass Filter Design". IEEE Transactions on Audio. v. AU-13, September/October, 1965, pp. 104-111.

15. Friend, D. "Resistance-Capacitance Active Filters for Audio Frequency Application". Journal of the Audio Engineering Society, v.16, January 1968, pp.94-110.

16. Fryer, W.D. "How to Design Low Cost Audio Filters". Electronics. v. 32, 10 April 1959, pp.68-70.

17. Gorski-Popiel, J. and A.J. Drew. "RC Active Ladder Networks". Proceedings of the IEE, v. 112, December, 1965, pp. 2213-2219.

18. Guillemin, D.A. Synthesis of Passive Networks. New York: Wiley, 1957.

19. Herro, J.L. and G. Willoner. Synthesis of Filters. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

20. Holt, B.F. Lost Cost Field Effect Transistor Measurements and Applications. MS Thesis. Monterey, California: Naval Postgraduate School, 1967.

21. Lewis, D. "Vocal Resonance." The Journal of the Acoustical Society of America. v. 8, October, 1936, pp.91-99.

22. Lindgren, N. "Directions for Speech Research." IEEE Spectrum. v. 5, March, 1968, pp. 83-92.

23. Lindgren, N. "Machine Recognition of Human Language." IEEE Spectrum. v. 2, March, 1965, pp.114-136.

24. Lindgren, N. "Speech -- Man's Natural Communication." IEEE Spectrum. v. 4, June, 1967, pp75-86.

25. Millman, J. and C.C. Halkias. Electronic Devices and Circuits. New York: McGraw-Hill, 1967, pp. 385-411.

26. Millman J. and H. Taub. Pulse, Digital, and Switching Waveforms. New York: McGraw-Hill, 1965, pp.317-319.

27. Mori, T. "Designing Low-Frequency Active Filters". Electro-Technology. v. 81, January, 1968, p.72.

28. Nichols, M.H. and L.L. Rauch. Radio Telemetry. Second Edition. New York: Wiley, 1956.

30. Peterson, G. E. and H.L. Barney. "Control Methods Used in a Study of the Vowels." The Journal of the Acoustical Society of America. v. 24, March, 1952, pp.175-184.

31. Phillips, N.D., Remillard, W., et al. "Teaching of Intonation to the Deaf by Visual Pattern Matching." Research Report RD-2360. Boston, Mass.: Northeastern University.

32. Potter, R.K., Kipp, G.A. and H.G. Kopp. Visible Speech. New York: Dover Publications, 1947.

33. Potter, R.K. and G.E. Peterson. "The Representation of Vowels and Their Movements." The Journal of the Acoustical Society of America. v. 20, July, 1948, pp.528-535.

34. Rosenstein, M. "A Vowel Meter and Its Application in Constrained Speech". IEEE Transactions on Audio. v. Au-13, November/December 1965, pp. 134-141.

35. Sallen, R.P. and E.L. Key. "A Practical Method of Designing RC Active Filters". IRE Transactions - Circuit Theory. v. 42, March , 1955, pp. 72-85.

36. Schroeder, M.R. "Vocoders: Analysis and Synthesis of Speech." Proceedings of the IEEE, v. 54, May 1966, pp. 720-734.

37. Shumard, C.C. "Design of High-Pass, Low-Pass and Band-Pass Filters Using R-C Networks and Direct Current Amplifiers with Feedback." RCA Review. v. 11, December, 1950, pp.534-564.

38. Sun, H.H. Synthesis of RC Networks. New York: Hayden, 1967.

39. Thiele, A.N. "Filters with Variable Cut-Off Frequencies." Institution of Radio and Electronics Engineers Australia v. 26, September, 1965, pp. 284-300.

40. Vogel, J.S. "Nonlinear Distortion and Mixing Processes in Field-Effect Transistors." Proceedings of the IEEE. v.55, December 1967, pp. 2109-2115.

41. Wallmark, J.T. "The Field-Effect Transistor -- A Review". RCA Review. v.23, December 1963, pp.641-660.

42. Williams, D.E. Visual Display of Certain Speech Parameters. MS Thesis. Monterey, California: Naval Postgraduate School, 1967.

43. Wise, C.M. Introduction to Phonetics. Inglewood Cliffs, N.J.: Prentice-Hall, 1958.

APPENDIX 1

Selected Glossary of Speech Terms

ARTICULATE. To produce a speech sound by the organs of speech.

ARTICULATION. The set of human bodily positions and movements aiming at the production of speech sounds.

BACK. A vowel articulated by raising the back part of the tongue towards the velum, e.g. sort.

CENTRAL. A vowel articulated by raising the central part of the tongue towards the juncture of the palate and the velum, e.g. first.

CONSONANT. A speech sound articulated by a complete closure of the air passage or by a narrowing of it beyond the vowel limit, e.g. go, or see.

DIPHTHONG. A vowel articulated by a deliberate movement of the speech organs from one position into the other.

FRICATIVE. A consonant articulated by a narrowing of the air-passage resulting in the audible friction, e.g. shame.

FRONT. A vowel articulated by raising the front part of the tongue towards the palate, e.g. get.

FULLY VOICED. A speech sound articulated by the vocal cords vibrating during the whole of its articulation, e.g. living or put.

ORGANS OF SPEECH. Those parts of the human body which are active in the production of speech sounds, i.e. the lungs the trachea (windpipe), the vocal cords, the glottis, the pharynx, the nose, the lips, the teeth, the alveoli (teeth ridge), the palate (hard palate), the velum (soft palate), the uvula, the tongue. The tongue is arbitrarily divided into four parts: the tip, the blade, the center and the back.

PHONEME. A class of distinctive speech sounds, the members of which are (1) in complementary distribution with each other, and (2) in opposition or contrast to any other class of distinctive speech sounds. Thus, /d/ in read and /d/ in middle are members of the same phoneme, whereas /d/ in date and /l/ in late are members of two different phonemes.

PHONEMICS. The scientific study of distinctive speech sounds.

PHONETICS. The scientific study of speech sounds.

PLOSIVE. A consonant articulated by a complete closure of the air passage, combined with air-compression behind the closure, and followed by an explosion in the release stage, e.g. kind.

SPEECH. A sequence of sounds articulated for the purpose of human communication.

SYLLABLE. A structural unit capable of being connected as a whole with one particular degree of accent, e.g. become.

VELUM. The soft palate of the oral cavity.

VOICED. A speech sound, consonant or vowel, articulated with the vocal cords vibrating during the whole of its articulation, or part of it, e.g. weather, park, one.

VOICELESS. A speech sound, especially a consonant, articulated with no voicing, e.g. lucky.

VOWEL. A speech sound articulated with no closure of the air-passage and no narrowing of it beyond the vowel limit, e.g. bad or most.

WORD. A structural unit separated in writing by spaces, e.g. bed (one word), room (one word), bedroom (one word), textbook (one word), a good subject (three words).

APPENDIX II

PHOTOGRAPHS OF PROTOTYPE EQUIPMENT

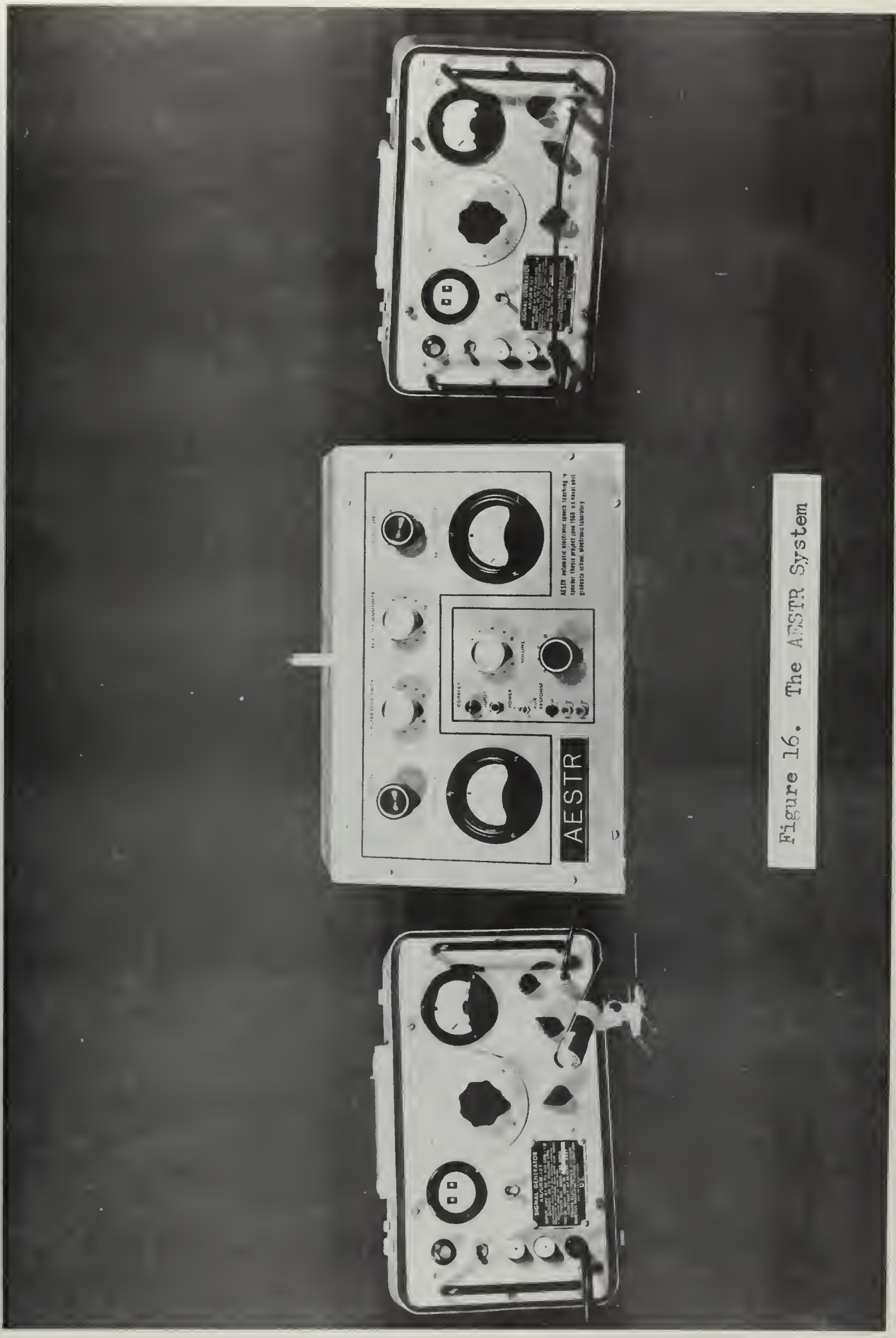


Figure 16. The AESTR System

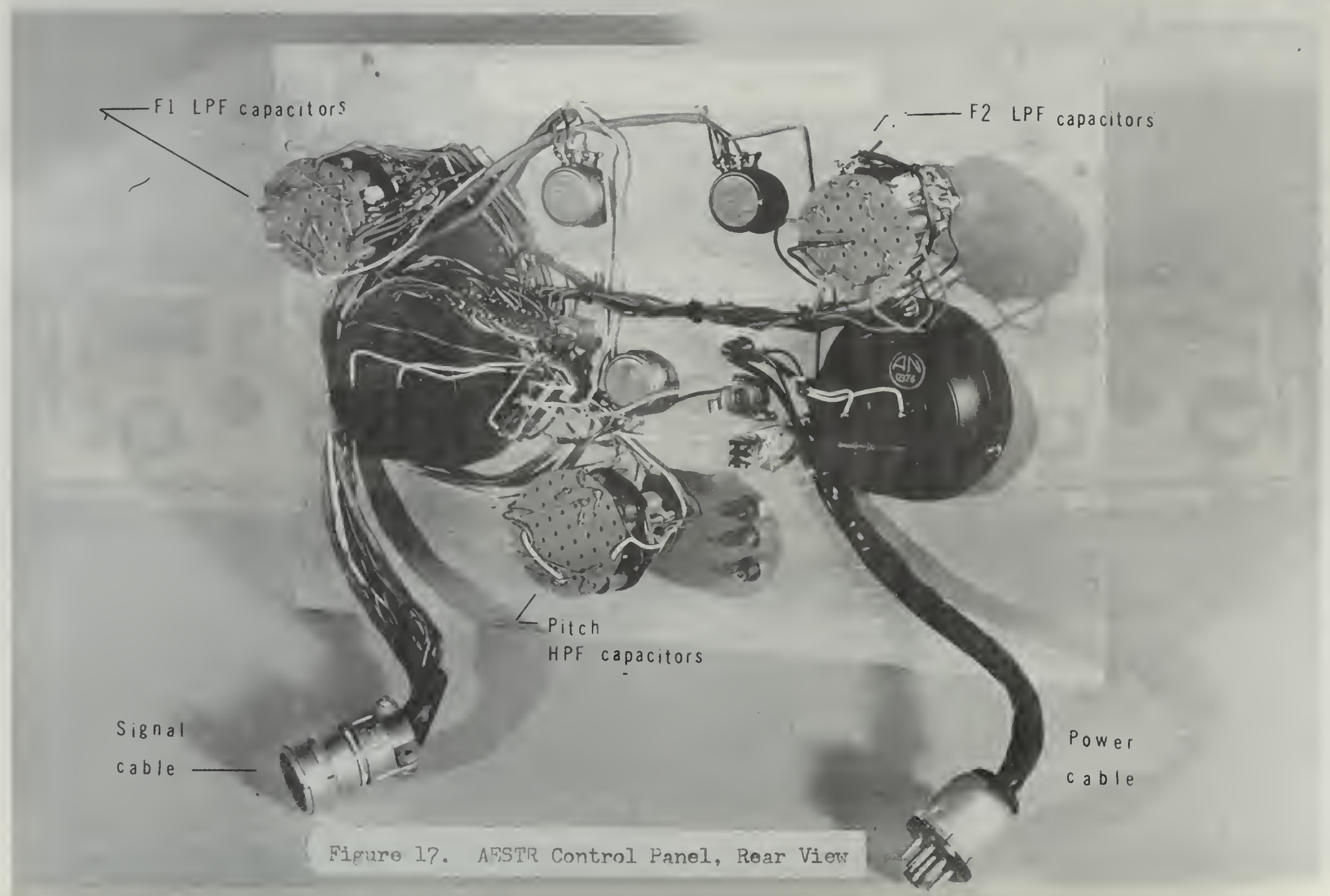


Figure 17. AFSTR Control Panel, Rear View

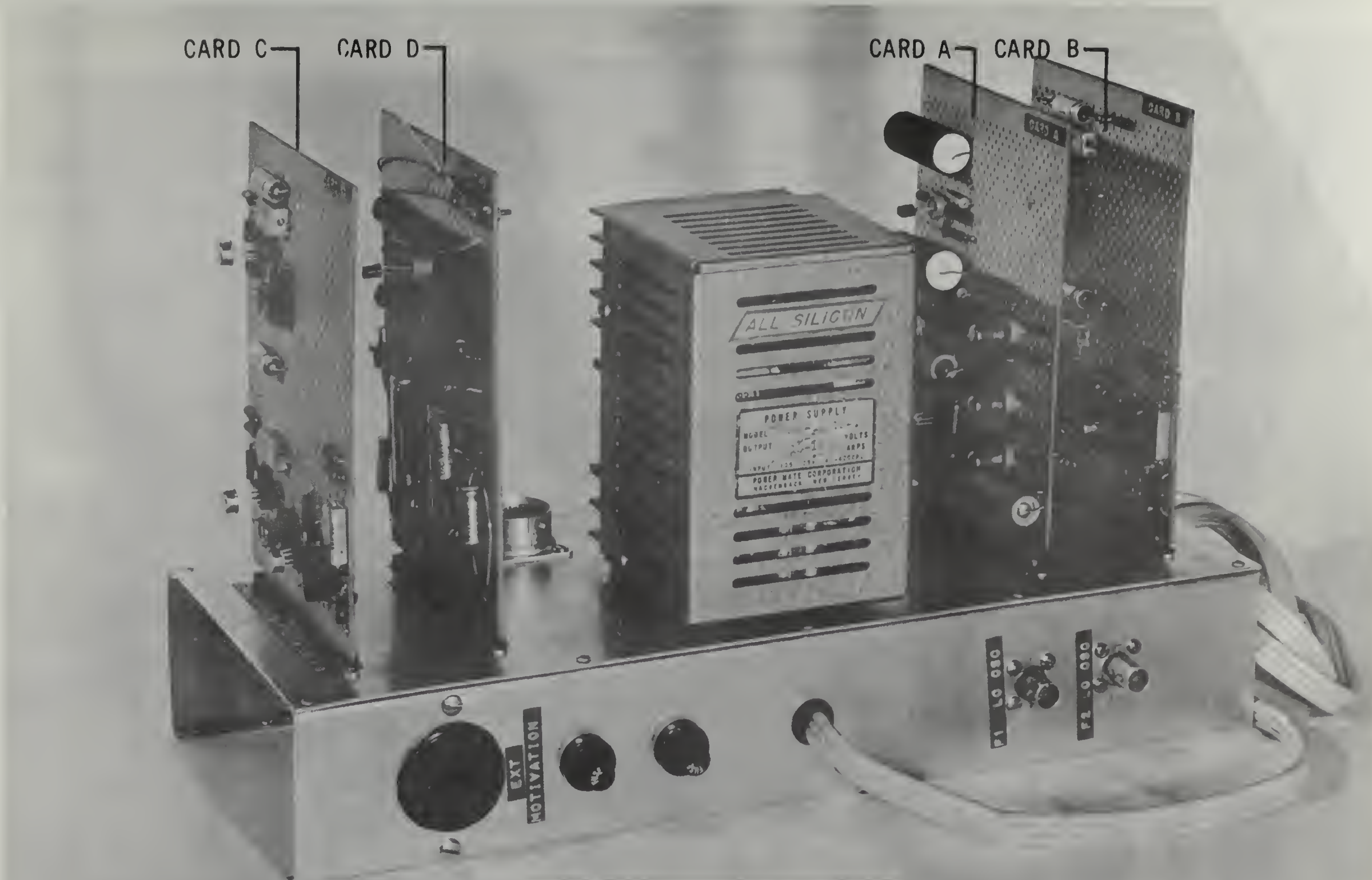


Figure 18. AESTR Chasis

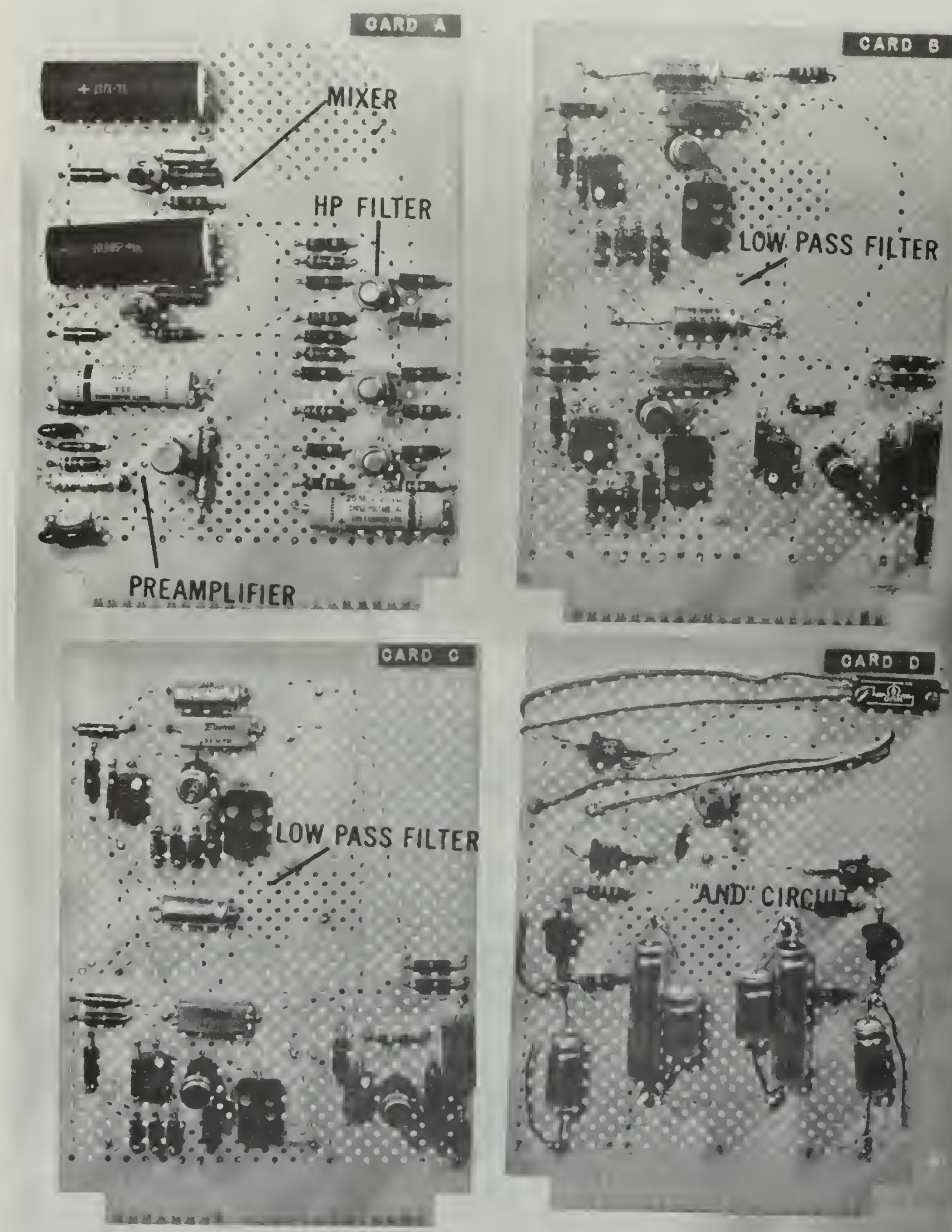


Figure 19. AESTR Circuit Boards

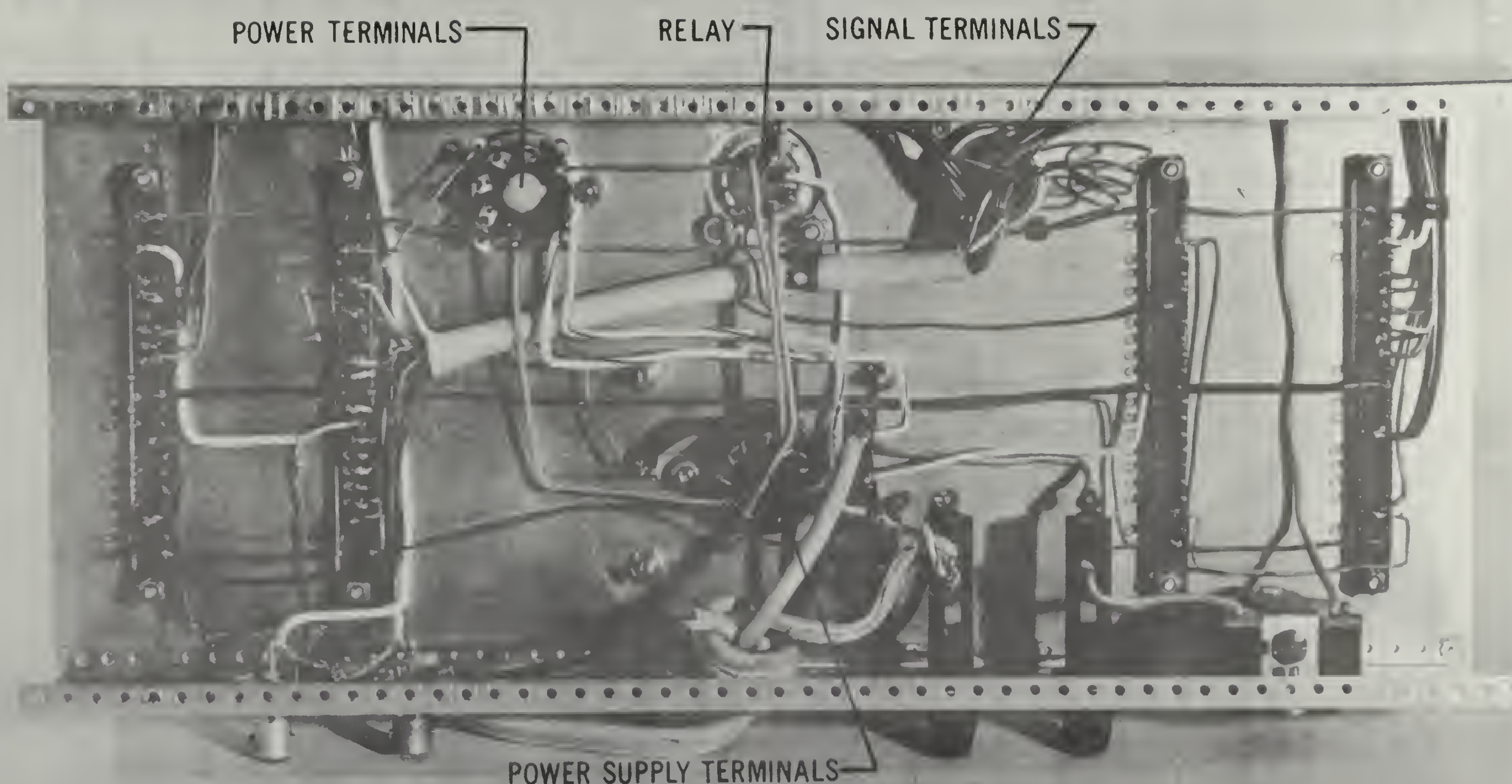


Figure 20. AESTR Chasis Internal Wiring

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
2.	Library Naval Postgraduate School Monterey, California 93940	2
3.	Director, U.S. Naval Security Group 3801 Nebraska Avenue, N.W. Washington, D.C. 20016	1
4.	Professor Gerald D. Ewing Department of Electrical Engineering Naval Postgraduate School Monterey, California 93940	4
5.	Academic Library Defense Language Institute West Coast Branch Monterey, California 93940	1
6.	LCDR Theodore J. Kuligowski USN 306 Alta Avenue Roseville, California 95678	1
7.	Mr. Victor McCullough Electronics Laboratory Naval Postgraduate School Monterey, California 93940	1
8.	Monterey Institute for Speech and Hearing 969 Pacific Street Monterey, California 93940	1
9.	Director of Naval Communications Department of the Navy Washington, D.C. 20350	1

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

An Automatic Electronic Speech Teaching Responder

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Thesis

5. AUTHOR(S) (First name, middle initial, last name)

Theodore J. Kuligowski
Lieutenant Commander, United States Navy

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1968	74	43

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d. Distribution unlimited	

10. DISTRIBUTION STATEMENT ~~This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior~~ approval.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

A concept for an apparatus which visually displays and responds to the first and second formant of vowel sounds is developed. The machine is intended for use by deaf and speech handicapped children in learning to produce voiced sounds. System design and principles applied to realize a physical prototype of this concept are presented. The complete electronic and mechanical design plus fabrication of the automatic electronic speech training responder is described in detail. Schematic diagrams of all electronic circuitry employed and photographs of the prototype equipment are included. The apparatus is on loan to the Monterey Institute for Speech and Hearing, Monterey, California for clinical testing and evaluation.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED

Security Classification

A-31408

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech						
Speech Display						
Speech Processing						
Speech Therapy						
Speech Training						

DD FORM 1 NOV 65 1473 (BACK)

S/N 0101-807-6821

UNCLASSIFIED

Security Classification

A-31409

~~NO FORM~~