presented at 103rd Conference of ACOUSTICAL SOCIETY OF AMERICA, April 1982

Noise Reduction in Speech Using Adaptive Filtering I: Signal Processing Algorithms

RICHARD W. CHRISTIANSEN, DOUGLAS M. CHABRIES,
and DOUGLAS LYNN

Abstract - **A study of the LMS adaptive filter as applied to the problem of noise suppression for speech signals with noise corruption is presented. Past work with LMS transversal filters is discussed. It is shown that the traditional techniques give rise to frequency distortion and the introduction of reverberation. An alternate frequency domain LMS adaptive filter is presented which circumvents the problems of distortion and reverberant echo formation. A tape demonstrating the concepts is played.**

Douglas M. Chabries and Richard W. Christiansen are on the faculty of the Electrical Engineering Department, Brigham Young University, Provo, Utah 84602.
Douglas Lynn is a student in the Electrical Engineering Department, Brigham Young University, Provo, Utah 84602

INTRODUCTION

Adaptive filters are filters that adjust themselves based on a given performance criteria. The most common of these is the LMS adaptive filter

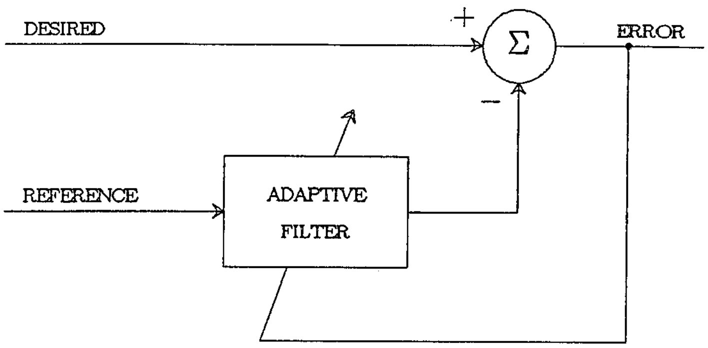


Figure 1. LMS Adaptive Filter

This structure, first proposed by Widrow in 1956, adjusts itself so as to minimize the mean square error between the desired input and the filter output [1 , 2]. By analyzing the expectation of signals at various points in the structure, it can be easily shown that components of the desired input that are correlated with components of the reference input will be cancelled from the error output leaving only uncorrelated components.

ADAPTIVE FILTER APPLICATIONS

There are two configurations of this structure commonly employed in the filtering of narrow band speech corrupted by noise. In the 'two microphone' or 'adaptive canceller' configuration, the speech and noise signal is presented to the desired input, while a sample of the noise alone is presented to the reference input.

Ideally the two noise inputs are correlated with each other while the speech and noise are uncorrelated. Thus, the noise components is removed from the signal, leaving speech in the error output. Any speech signal present at the reference input limits the

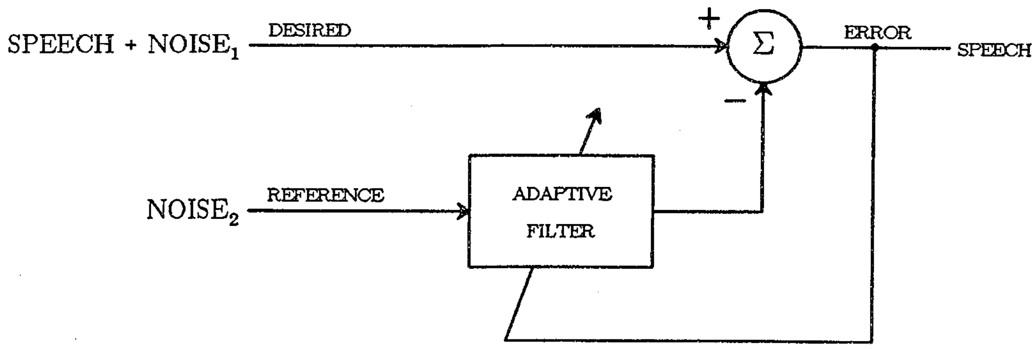


Figure 2. LMS Adaptive Noise Canceller

maximum possible signal to noise gain to the inverse of the speech to noise ratio at the reference input.

In applications where an independent sample of the noise is not available, the 'one microphone' or 'adaptive enhancer' configuration is employed [3].

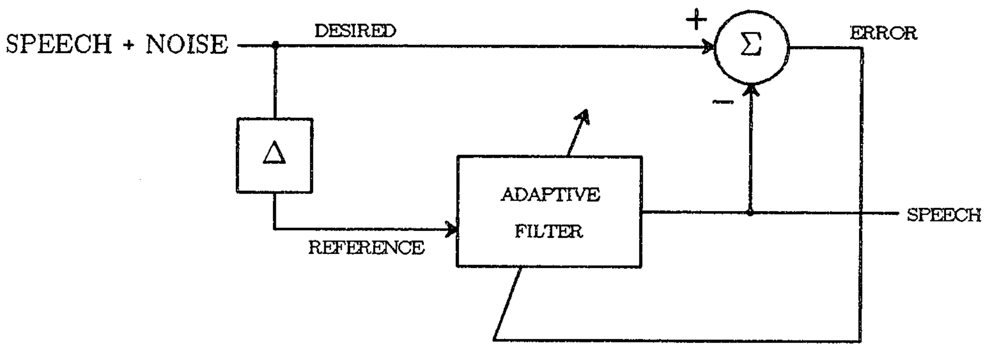


Figure 3. LMS Adaptive Line Enhancer

In this configuration, proposed by Widrow in his original paper and later by Sambur [4], a delayed version of the speech and noise applied to the desired input is also applied to the reference input. The delay is chosen such that the noise components of the desired and reference inputs are uncorrelated with each other while the signal components remain correlated. As before, to minimize the mean square error, the correlated components, in this case the speech, are cancelled leaving noise in the error output, and speech in the filter output.

To explain some of the constraints on the use of adaptive filters for cancelling noise from speech, we examine the filter structure in more detail.

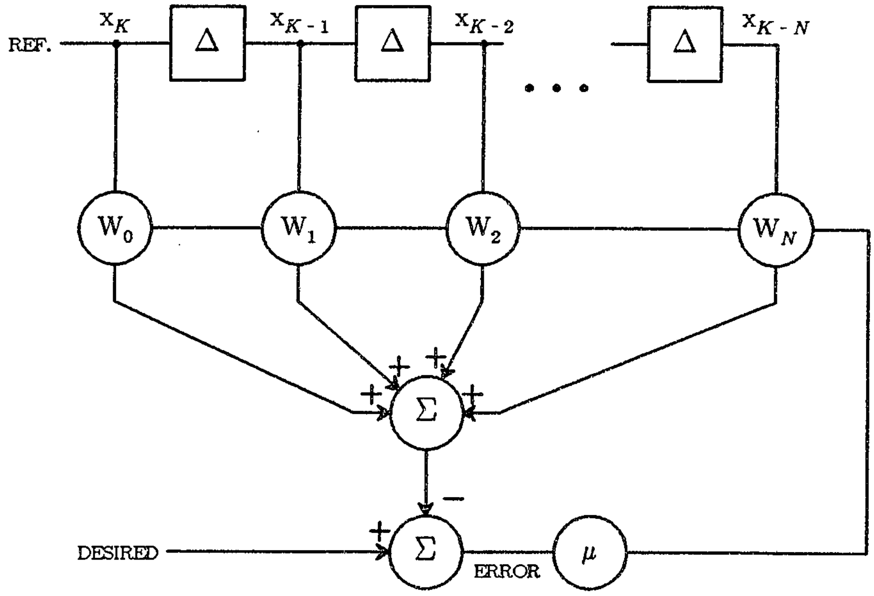


Figure 4. Transversal Adaptive Filter Structure

The structure itself is the standard tapped delay line filter except that the error output is used to modify the filter weights.

The relationship between the mean square error and the weight values is quadratic. A plot of the mean square error against a single weight yields a parabola. Plotting the mean square error against N weights in N dimensions yields a concave hyperparaboloidal surface. To minimize the mean square error, the weights are adjusted according to the negative gradient of this error surface.

The weight update consists of computing an estimate of the gradient; scaling it by an adaptive learning constant, μ ; and subtracting this from the previous weight value.

If μ is chosen too large, the weight may skip back and forth across the error bowl either actually increasing the time required to reach the bottom of the bowl, or causing the weight to climb up to sides of the bowl, eventually causing the filter to blow up. If μ is chosen too small, the weight may never reach the bottom of the bowl. Both these effects become even more marked when the statistics of the data, and hence the bowl position and shape are changing with time. Where speech is concerned, the vocal tract is stable for periods between 10 to 30 msec. So an enhancer would have to adapt at

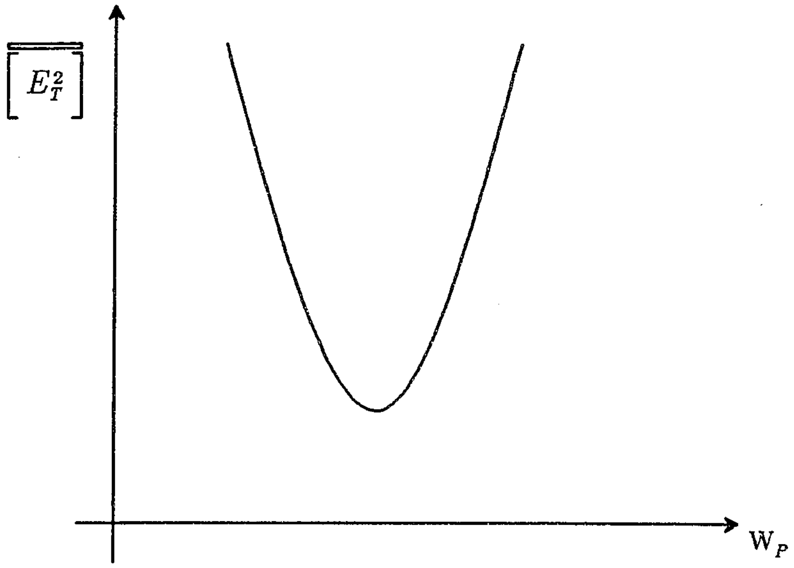


Figure 5. Quadratic Surface for Mean Square Error

least that fast. Since the canceller operates by learning the nature of the correlation of the noise inputs, it must only adapt as fast as the correlation of the noise inputs may be changing. This correlation is most often related to room acoustics.

Another, and perhaps at first less obvious problem is related to speech statistics. Wide spread eigenvalues in the speech and noise autocorrelation matrix give rise to skewed mean square error surfaces. Moving along the gradient on such an error surface causes the weights to approach their final values at different rates. What this means is that components of the signal that have the most power and thus contribute most to the skewness of the error surface are resolved first, leaving lower power components to be resolved later. In speech it is the lower frequency components that carry most of the energy hence the filter quickly captures the low frequency portion of speech, leaving the high frequencies to be resolved much more slowly. Since it has been shown that the high frequencies carry much of the intelligibility of speech, this effect is highly undesirable. On the average, this results in a high frequency de-emphasis that subjectively "muffles" the speech.

FREQUENCY DOMAIN LMS FILTER

We at BYU have proposed two possible solutions to this problem which are found in the context of a frequency domain adaptive filter. These solutions follow in the footsteps of Dentino [5], but utilize additional degrees of freedom and compensate for circular convolution effects. The first is employing weight update algorithms that adapt more quickly to the final solution, thus rendering the effect negligible. The second involves transforming the weightspace so that any skewness in the error surface now lies parallel with the dimensions of the new weightspace and then employing a separate μ learning constant in each dimension thus normalizing the adaptation over the various dimensions.

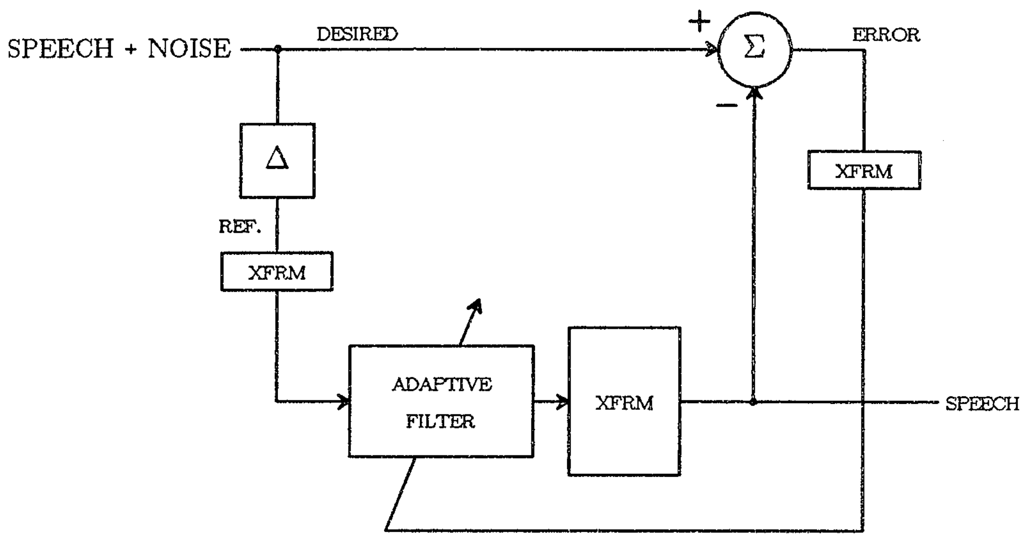


Figure 6. Block Diagram for the Frequency-Domain
Normalized Weightspace Filter

Care must be take to avoid the effects of circular convolution. This is typically done using either the overlap and add technique or the overlap and save technique. The overlap and save technique is the most advantageous in this particular application. Selection of a separate μ for each component of the new weightspace allows the user to control the adaptation for each bin of the adaptive filter, thereby eliminating the muffling and reverberation characteristic of the time domain approaches.

Results of the foregoing implementations are presented to demonstrate the effects.

CONCLUSION

The application of adaptive filters to the problem of noise suppression given that no independent sample of the noise is available has received considerable attention in the literature. The traditional time-domain techniques result in both muffling and reverberant distortion. A frequency domain technique is described, which in this application removes these effects. Data is presented to substantiate the foregoing claims.

References

1. B. Widrow, "Adaptive Filters," in *Aspects of Network and System Thoery*, Rinehart and Winston, New York, 1970.
2. B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. C. Goodlin, "Adaptive Noise Cancelling: Principles and Applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692-1716, Dec. 1975.
3. J. R. Zeidler, E. H. Satorius, D. M. Chabries, and H. T. Wexler, "Adaptive Enhancement of Multiple Sinusoids in Uncorrelated Noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 3, June 1978.
4. M. R. Sambur, "Adaptive Noise Cancelling for Speech Signals," *IEEE Trans. ASSP*, vol. ASSP-26, no. 5, pp. 419-423, October 1978.
5. M. Dentino, J. McCool, and B. Widrow, "Adaptive Filtering in the Frequency Domain," *Proceedings IEEE*, vol. 66, pp. 1658-1659, Dec. 1978.

NOISE REDUCTION IN SPEECH USING ADAPTIVE FILTERING I:
SIGNAL PROCESSING ALGORITHMS

PAPER PRESENTED BY:

R. W. Christiansen
D. M. Chabries
D. Lynn

AT THE 103RD MEETING OF THE
ACOUSTICAL SOCIETY OF AMERICA

April, 1982
Chicago, Illinois

NORTHWESTERN UNIVERSITY
THE SCHOOL OF SPEECH
EVANSTON, ILLINOIS 60201

HUMAN COMMUNICATION SCIENCES
DEPARTMENT OF COMMUNICATIVE DISORDERS
FRANCES SEARLE BUILDING
2299 SHERIDAN ROAD

1 February 1982

From: Technical Program Chairperson

To: R. W. Christiansen

Your paper, Noise reduction in speech using - - - -

has been accepted for presentation at the Chicago meeting of the Acoustical Society of America on April 27, 1982 in Session C at 9:00 A.M. P.M. Your presentation will be in the (lecture, poster) type format for which you have been allocated 12 minutes. If your paper has co-authors, please forward this information to them since this notification is being sent only to the person whose name was listed at the bottom of the abstract.

Please contact your Session Chairperson at the Chicago meeting prior to the beginning of the session to let him or her know you are there and to receive last minute information. Instructions are enclosed and should be followed closely; do not hesitate to call me (312 492-3180), in advance of the meeting, if you have any questions.

Sincerely,

Lois Elliott

Lois L. Elliott
Technical Program Chairperson

Noise reduction in speech using adaptive filtering II: Speech intelligibility improvement with normals and hearing impaired subjects. R. H. Brey (Communicative Disorders Area, Brigham Young University, Provo, Utah 84602) and M. S. Robinette (University of Utah, Department of Communication)

An adaptive LMS filter is employed to filter speech in signal-to-noise ratios varying from -8 dB. to +12 dB. The filter configuration used is that previously proposed by Widrow, Boll, and Pulsipher and commonly called noise cancellation. Intelligibility measures were obtained using speech selected from (CID) W-22 word lists. Fifteen lists of 50 words each were selected. Ten normal hearing and six hearing impaired individuals were subjects. Sound field and free field measurements, as well as processed and unprocessed data were obtained and analyzed.

Intelligibility was near zero for all SNR's less than 0 dB prior to processing. Following processing scores increased 30 to 40% for both normal and hearing impaired groups. For SNR's of +4 dB to +12 dB, intelligibility scores prior to filtering ranged from 10% to 50%; after filtering the scores improved to within 10% of the noise-free sound-field measurement. The specific test configuration, SNR improvement estimates, as well as additional results and comparisons will be presented.

Technical Committee: Speech Communication

Subject classification number(s): 43.70.Qa use of computers in speech studies.

Telephone Number: (801) 378-4319 (R. H. Brey)

Send acceptance or rejection notice to R. H. Brey

NOISE REDUCTION IN SPEECH USING ADAPTIVE FILTERING II: SPEECH INTELLIGIBILITY IMPROVEMENT WITH NORMAL AND HEARING IMPAIRED STUDENTS

PAPER PRESENTED BY:

R. H. Brey
M. S. Robinette

AT THE 103RD MEETING OF THE
ACOUSTICAL SOCIETY OF AMERICA

April, 1982
Chicago, Illinois

A major complaint of those with hearing impairments is a reduced ability to understand speech in a noisy environment. Even without hearing impairment, the addition of background noise may significantly reduce speech intelligiblity.

The previous paper presented an adaptive canceller scheme which can be used to reduce the noise when a second correlated sample of noise is available.

This paper treats the adaptive canceller implemented in the time domain. Speech intelligibility was measured in several signal to noise ratios, with and without adaptive filtering, using both normal and hearing impaired subjects.

A room approximately 10x15 feet was used to make recordings with Speech Spectrum Noise played through a loudspeaker. This configuration provided an interference for noise cancellation in a realistic environment where the interference is composed of multiple reflections of the noise off the walls and equipment in the room. The length of the adaptive filter was set at 688 taps with a 20 KHz sample rate.

Processing was performed in real time. A small time delay (48 samples = 2.4 ms) was inserted into the interference reference channel of the adaptive noise canceller to account for propagation delay. The speech materials chosen for the intelligibility tests were the (CID) W-22 phonetically balanced word lists.

Figure 1

The equipment was arranged such that the noise was fed directly into the reference input of the adaptive filter. This same noise signal was then routed to the left speaker. The W-22 lists were played from a tape recorder through the right speaker. Either the speech, or the speech plus the noise were picked up by the microphone placed 150 cm equidistant from both speakers. The microphone output was then recorded directly onto channel two of another tape recorder and also routed to the primary input of the adaptive filter. The output of the adaptive filter was then recorded on channel one of the same cassette tape recorder.

Reverberation measurements in the test room ranged from .317 to .404 seconds for 250 to 4,000 Hz using 1/3 octave narrow band noises. SNR's were from -8 dB to +12 dB in increments of 4 dB.

Subjects and procedures

Twelve normal hearing subjects (mean age 23 years) were required to obtain a score of 90-100% on word list 1A from a direct tape under earphones. All subsequent test tapes were administered at 60 dB HTL to the right ear of the normal hearing subjects.

The subjects were played either the processed or unprocessed tapes at the various SNRs or in quiet in a counterbalanced order. The lists were always presented from the poorest SNR in increasing 4 dB increments to the highest SNR to avoid learning effect from listening to different randomizations of lists 3 and 4.

Data obtained on six hearing impaired subjects were collected in the same manner except that the test tapes were played at a Most Comfortable Listening Level.

Figure 2

Results Normal Hearing Subjects

The histogram portion in Fig. 2 shows the twelve normal hearing subjects obtained a mean score of 99.5% on the initial screening list (the recording not contaminated by room acoustics). Their intelligibility scores using the tapes recorded in the sound-field without noise present were 82.5% for the unprocessed and 82% for the processed recordings. Therefore all subjects scored normally on the tapes recorded directly from the master record but dropped approximately 18% when they were retested using tapes contaminated by the acoustic characteristics of the sound-field environment.

Of primary importance, however, are the other data shown in this figure. The lower curve shows the mean data for the various SNRs in the unprocessed condition. Note that the mean scores at -8, -4, and 0 dB SNRs were all 4% or less. The curve then begins to rise to 25, 42, and 46% correct. Error bars represent plus and minus one standard deviation (SD).

The upper curve represents the same SNRs after the adaptive filtering process. Note, for the scores at -8, -4, and 0 dB SNRs wherein it was difficult to recognize the presence of speech in the unprocessed condition, subjects obtained mean correct scores of 31, 46, and 50% respectively. The curve then continues to rise to 75% correct at the +12 dB SNR.

Although the maximum mean score of 75% appears low, remember that the maximum mean score in the sound-field condition without noise present was 82.5%. In fact, the error bars of 1 SD overlap for both conditions at the +8 and +12 SNRs.

D E M O N S T R A T I O N   T A P E

The demonstration tape FIRST presents two CID words from the Master tape.

T A P E

Next, two words from the sound-field unprocessed tape, followed by two words processed through the adaptive filter without noise present.

T A P E

Note the 3 KHz equipment artifact present in the equipment prototype. Next, three words are presented at -8 dB SNR unprocessed, followed by two words processed.

T A P E

Next, two words are presented at 0 dB SNR unprocessed, followed by two words processed.

T A P E

Finally two words are presented at +8 dB SNR unprocessed, followed by two words processed.

Results Hearing Impaired Subjects

Figure 3

Fig. 3 shows the group audiogram of the right ear of five adults with a sloping high frequency sensorineural type hearing loss.

Figure 4

Fig. 4 shows the group data of intelligiblity scores, processed and unprocessed, across the six signal to noise ratios. Note that the five hearing loss subjects show scores surprisingly similar to the normal hearing group.

Figure 5

Fig. 5 shows this similarity by superimposing the normal hearing, mean data over the data of the five hearing impaired subjects.

Figure 6

Fig. 6 shows the audiogram of a sixth hearing impaired subject who has a low frequency sensorineural type hearing loss. Although his hearing shows a different loss by frequency configuration, his processed vs unprocessed speech intelligibility shows a similar pattern to the other five hearing impaired, as well as the normal hearing group.

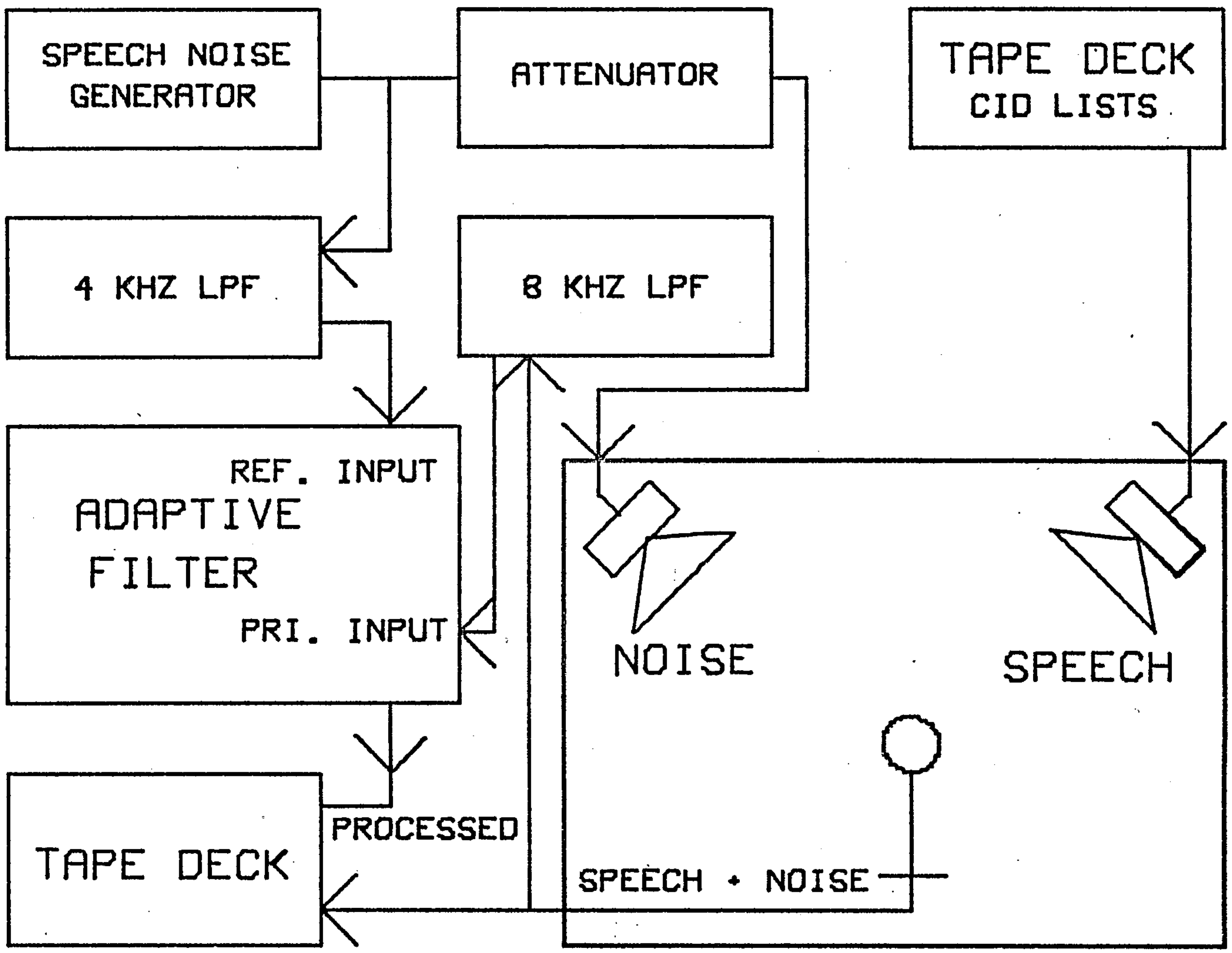
Figure 7

Fig. 7 emphasizes this similarity by superimposing the mean normal hearing subject data on the processed vs unprocessed scores of this low frequency hearing loss subject.

The data described definitely point out the advantage of a noise cancellation system for a situation wherein the possibility exists, of obtaining a reference sample of the noise, independently from the primary speech signal mixed with the noise. Such conditions exist in many situations. The authors are presently undertaking plans to employ this system in an industrial high noise environment.

T H A N K Y O U

BYU



EQUIPMENT DIAGRAM

Fig. 1

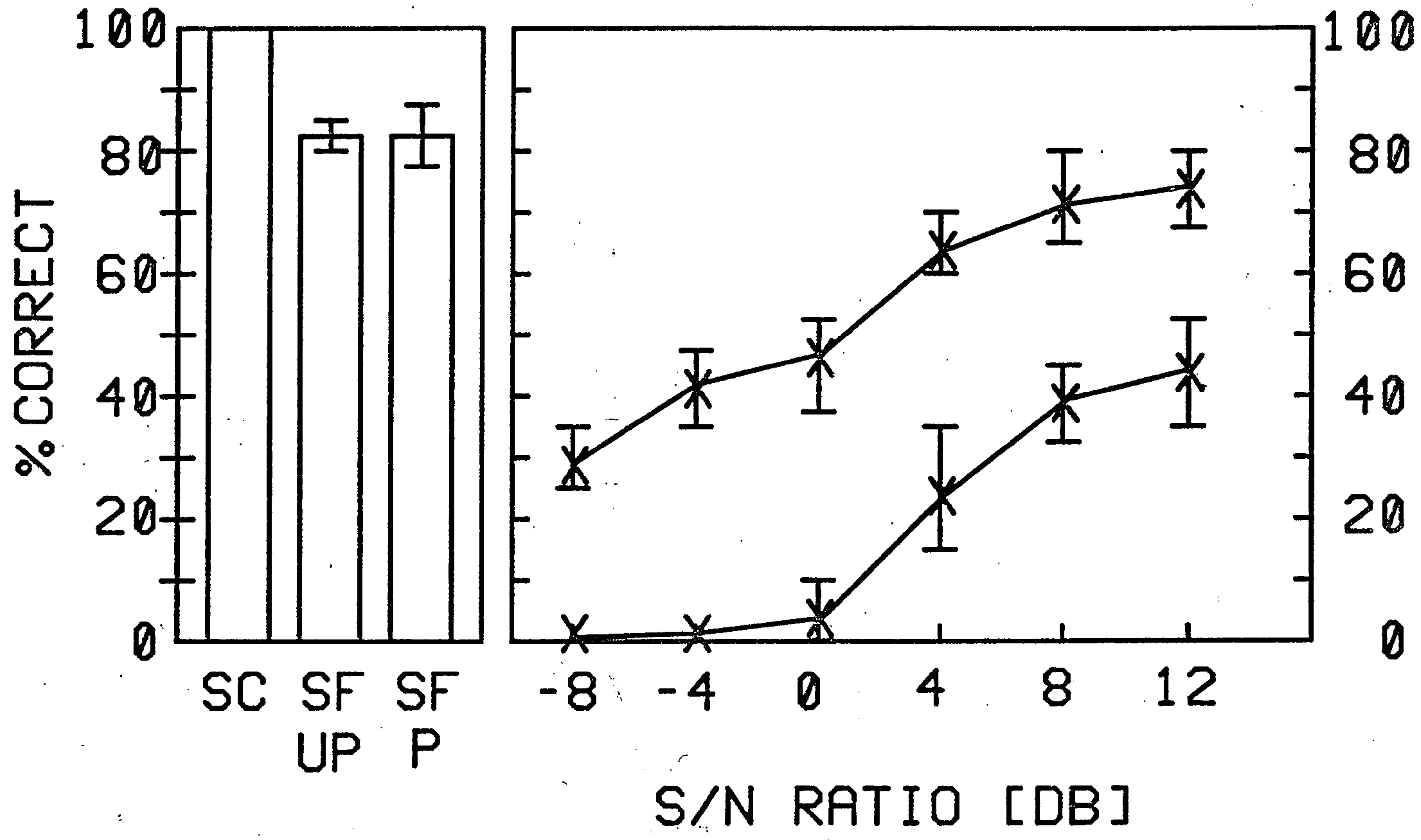


Fig. 2

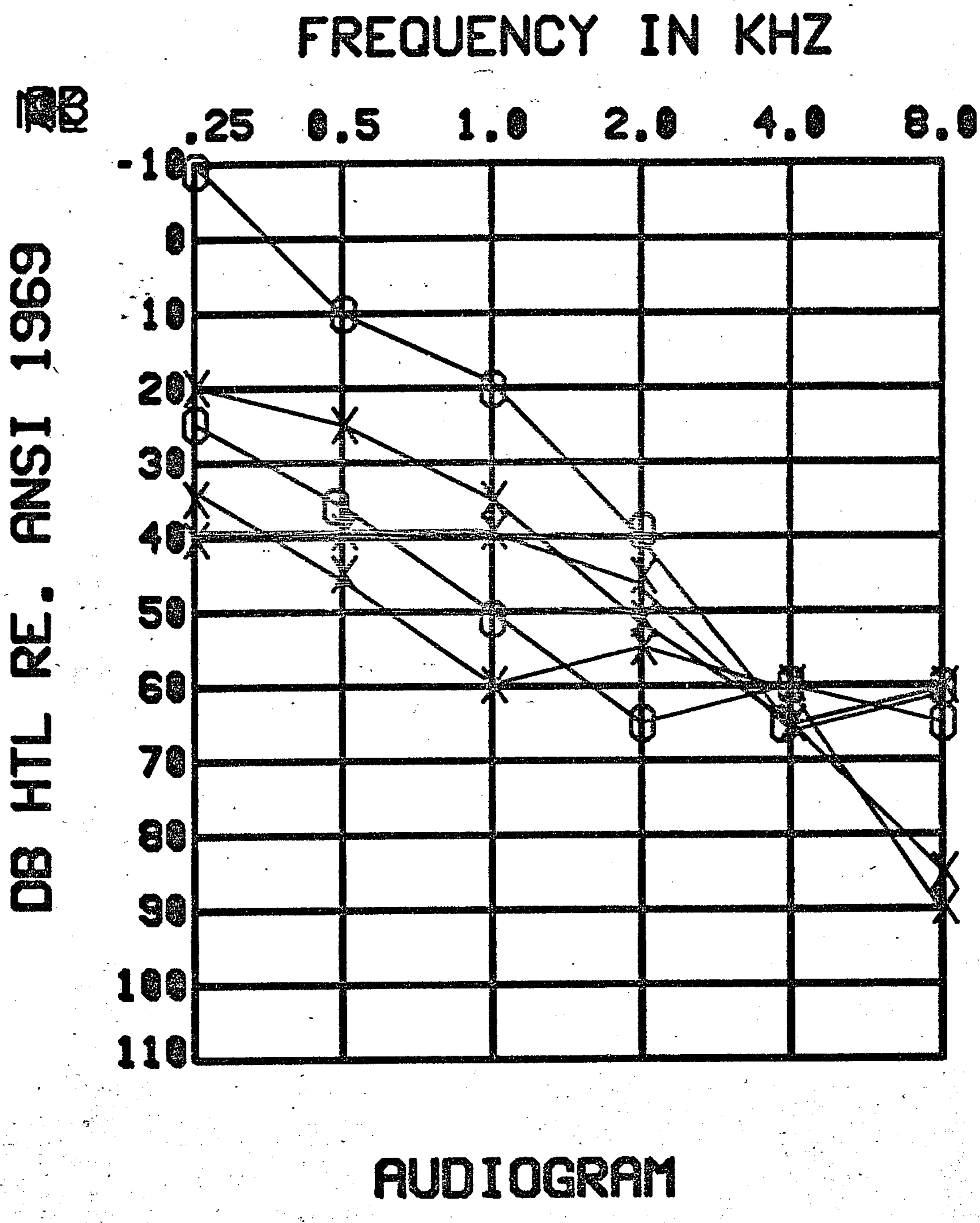
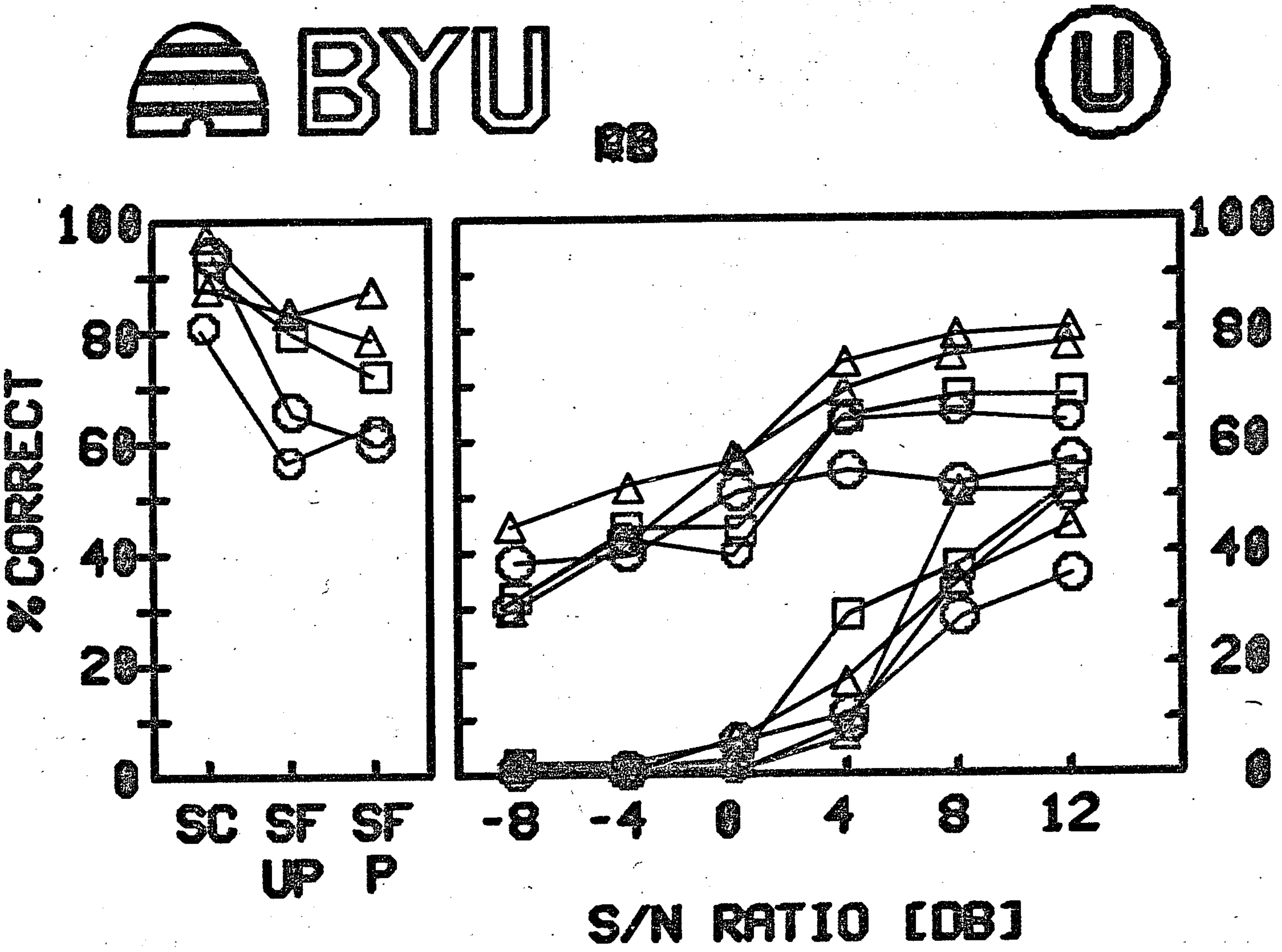

FREQUENCY IN KHZ
DB
.25
0.5
1.0
2.0
4.0
8.0
-10
0
10
20
30
40
50
60
70
80
90
100
110
DB HTL RE. ANSI 1969
AUDIOGRAM

Fig.3



BYU
RB
U
% CORRECT
100
80
60
40
20
0
SC
SF UP
SF P
-8
-4
0
4
8
12
S/N RATIO [DB]

Fig. 4

BYU RB

U

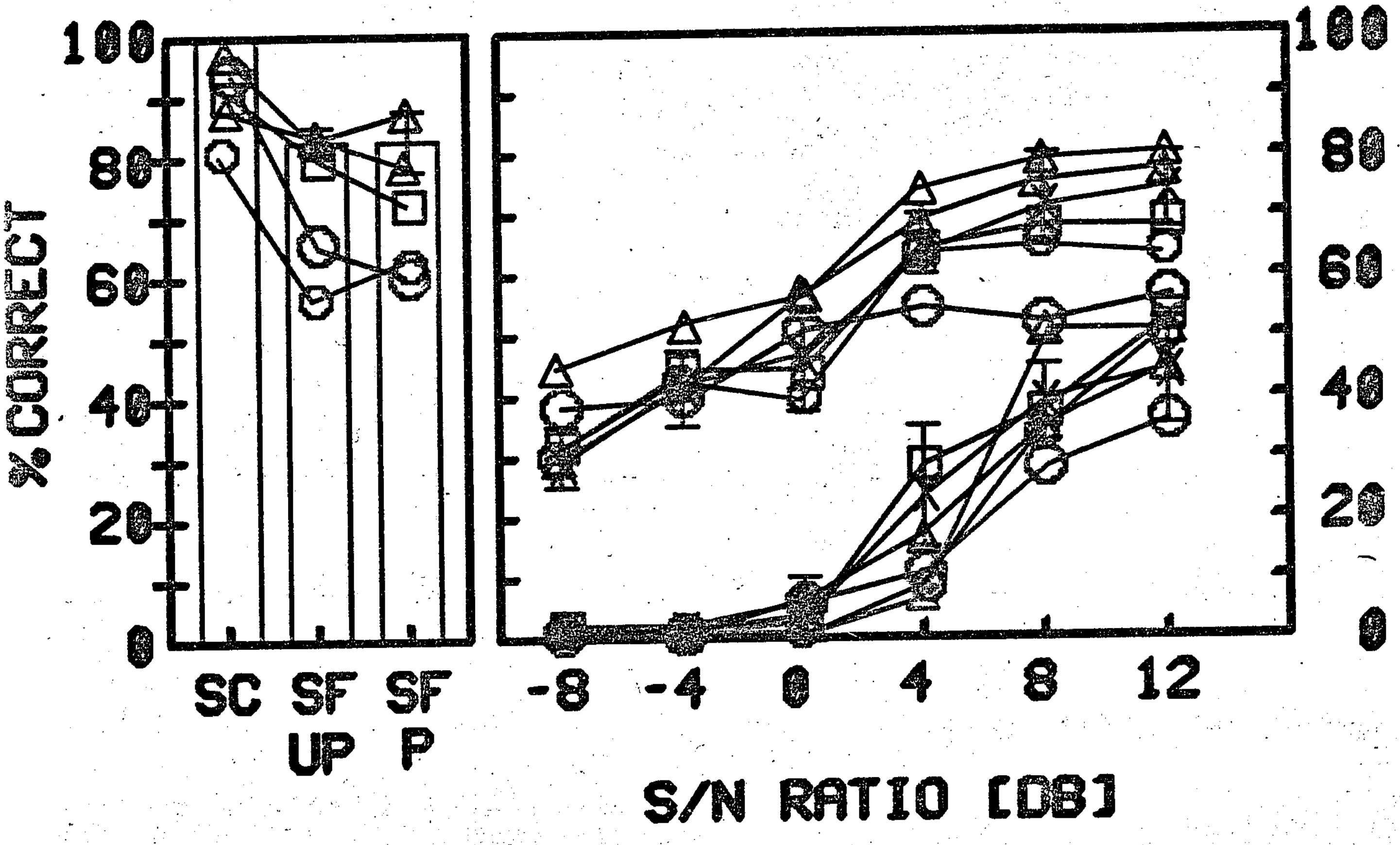


Fig. 5

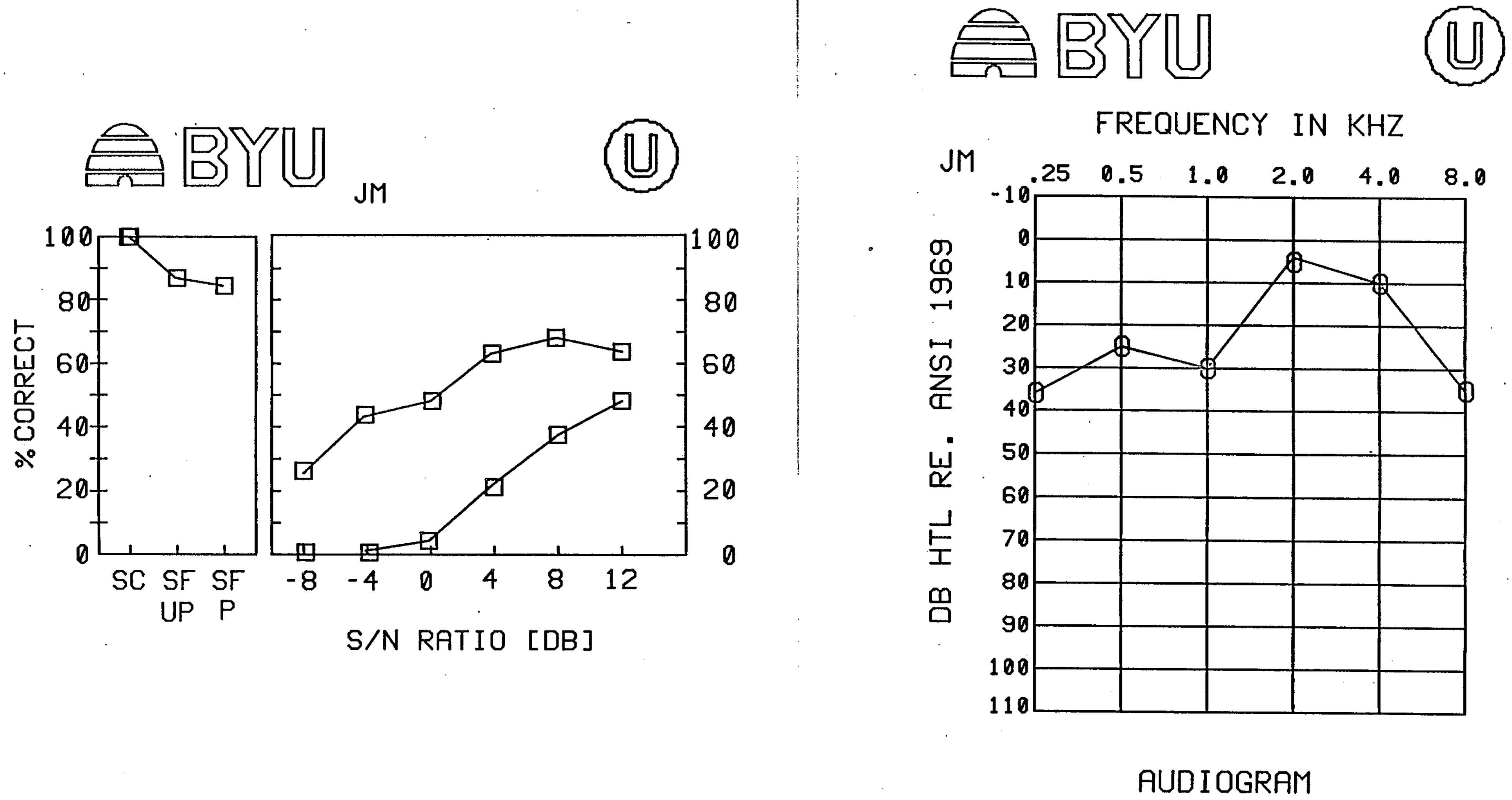
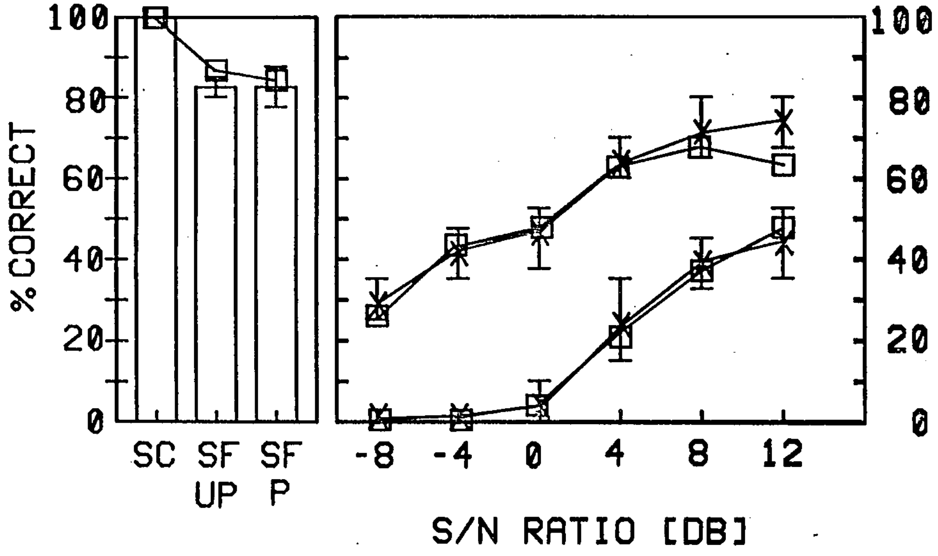

BYU
JM
% CORRECT
SC SF SF
UP P
S/N RATIO [DB]
BYU
FREQUENCY IN KHZ
JM
.25 0.5 1.0 2.0 4.0 8.0
DB HTL RE. ANSI 1969
AUDIOGRAM

Fig. 6


JM
% CORRECT
100
80
60
40
20
0
SC
SF UP
SF P
-8
-4
0
4
8
12
S/N RATIO [DB]

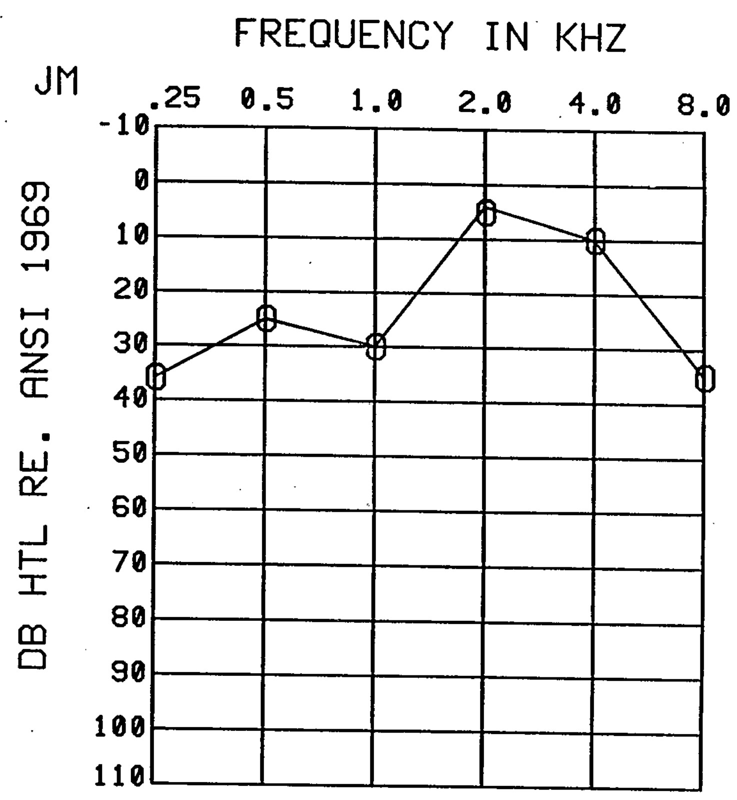

FREQUENCY IN KHZ
JM
.25
0.5
1.0
2.0
4.0
8.0
DB HTL RE. ANSI 1969
-10
0
10
20
30
40
50
60
70
80
90
100
110
AUDIOGRAM

Fig. 7