NORTHWESTERN UNIVERSITY

DESIGN AND EVALUATION OF HIGH-FIDELITY HEARING AIDS

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Audiology

By

MEAD CLIFFORD KILLION

Evanston, Illinois

June, 1979

ABSTRACT

Killion, Mead Clifford (Ph.D. Field of Audiology)

DESIGN AND EVALUATION OF HIGH-FIDELITY HEARING AIDS

Directed by Tom W. Tillman

(Words in Abstract: 500)

While recent research data indicate a consensus may have been
reached on the question of how to maximize speech discrimination with
some types of hearing loss, very little data are available on the
question of just what it is that should be optimized in order to max-
imize user satisfaction with a hearing aid. The task of determining
the electroacoustic characteristics of a hearing aid appropriate to a
given type of hearing loss is complicated by a wide variety of factors.
For the user with a mild-to-moderate loss, "aided speech discrimination"
may be one of the less important factors. This problem is discussed at
some length in the first chapter.

If a mild-to-moderate cochlear hearing loss which consists of
nothing more than a loss of sensitivity for low-level sounds is as-
sumed, however, it becomes possible to infer the optimum hearing aid
for that loss: The hearing aid should provide gain for low-level
sounds but do absolutely nothing (subjectively disappear) for high-
level sounds. At high levels, therefore, such a hearing aid would be
nothing more or less than a unity-gain high-fidelity sound-reproduction
system as judged by someone with normal hearing (or by someone with
the normal high-level hearing implied by the assumption above).

Guidelines for such high-fidelity hearing aid design can be de-
duced from known objective and subjective factors. These factors
include head diffraction and external-ear resonance, the bandwidth and
peak levels required to reproduce live (e.g., orchestral) performances,
the nonlinear distortion characteristics of earphones and ears, the
masking effect of common background noise levels, etc. In all cases,
existing hearing-aid transducer and amplifier technology can be shown
adequate to those guidelines. These issues are discussed in a second
chapter, along with a description of two pairs of laboratory hearing
aids constructed in accordance with such guidelines. One is a prac-
tical Over-the-Ear design with 8-kHz bandwidth, while the other is an
In-the-Ear design with 16-kHz bandwidth. Frequency response, distor-
tion, and overload measurements performed on those hearing aids indi-
cated their objective performance placed them in the high-fidelity
class.

In order to obtain subjective fidelity ratings of the hearing
aids, they were rated by three subject groups on a simulated live-
versus-recorded A-B-A comparison task along with several common high-
fidelity loudspeakers and headphones. The experimental design and
results are described in the third and fourth chapters. The fidelity
ratings given the experimental hearing aids by all three subject groups
(Untrained Listeners, Golden Ears, and Trained Listeners) placed them in
the Good-to-Excellent fidelity class. The hearing aid ratings were

similar to those given to very-high-quality sound systems and signifi-
cantly higher than those given moderate-quality sound systems such as
a speech audiometer.

These results support the following conclusion: The important
question for hearing aid research is no longer "What can a hearing aid
be designed to do?", but "What should a hearing aid be designed to do
for the hearing impaired?"

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

A REVIEW OF "THE HEARING AID PROBLEM"

INTRODUCTION

People have been complaining about hearing aids for years. Audiologists, hearing aid wearers and others have all suggested at one time or another that the "problem" with hearing aids was: (a) their poor electroacoustic characteristics (limited frequency response, high distortion, etc.) and (b) an inferior delivery system which was generally incapable of matching the right device with the right person.

These issues take on added significance now that previous limitations on subminiature transducer performance have been largely removed. The frequency response of currently available subminiature microphones can be tailored to cover the range from a sharply rising bandpass filter to a wideband response which is flat within a few dB from 20 Hz to 20 kHz (Killion and Carlson, 1970, 1974). Their noise level can be reduced to the order of that of good young ears (Killion, 1976a). Furthermore, wideband earphones (Carlson et al., 1976) in conjunction with appropriate earmold plumbing (Carlson, 1974, Killion, 1976b) can produce a smooth frequency response from 20 Hz to 8 kHz or more with undistorted sound output equivalent to 110-115 dB free-field SPL.

Concurrent with the removal of previous transducer limitations, advances in solid state technology have made extensive signal processing (stimulus-response changes in the amplitude, frequency, and/or time domain) available in extremely compact devices. The potentials of both analog and digital signal processing have barely been tapped to date.

Thus, the question is no longer "What can be done?" but "What should be done?" in hearing aid design. One possibility is that a high fidelity hearing aid will prove useful. A demonstration that high-fidelity hearing aids are now practical is a major component of this thesis.

Consider, however, that the term high fidelity itself may have a different meaning for someone with a hearing impairment than for someone with normal hearing. Briefly, we can think of a high-fidelity hearing aid as consisting of a sound reproduction system which is high fidelity in the conventional sense, in combination with one or more forms of low-distortion signal processing acting to compensate for the hearing impairment. In the case of a pure conductive hearing loss, for example, such signal processing might consist of nothing more than a linear amplifier with gain equal to the hearing loss. The signal processing required to even partially compensate for a severe frequency-dependent sensorineural hearing loss might be quite complex, on the other hand. Our primary interest here is in the individual with mild-to-moderate hearing loss who has no difficulty discriminating speech, enjoying full-frequency-range sound reproduction, etc., provided only that the sound intensity is great enough. (We thus exclude those users whose hearing loss at some frequencies is so severe that high-fidelity sound reproduction in the conventional sense would have little meaning.)

Determining the optimum signal processing for such a loss is beyond the scope of the present study, although several essential features can be deduced if certain simplifying assumptions are made about the characteristics of a "typical" hearing loss. In any case, a wearable sound reproduction system which is high fidelity in the conventional sense is an essential building block for a truly high fidelity hearing aid. That building block is the topic of the present study.

The organization of this dissertation is as follows: A discussion of some of the many factors which must be included in the overall "goodness rating" of a hearing aid is contained in this Chapter, which is divided into three sections. The first section contains a review of (frequency) selective amplification versus speech discrimination; i.e., the effect of signal processing in the frequency domain on speech discrimination. The second section discusses several complicating factors which arise when attempting to evaluate any hearing aid (factors other than speech discrimination). The third section contains a brief discussion of the rationale for a high-fidelity hearing aid.

Chapter II contains a discussion of several of the design considerations appropriate to high-fidelity hearing aids, such as bandwidth, response smoothness, nonlinear distortion, input dynamic range, etc.; as well as a discussion of some of the problems peculiar to a high-fidelity hearing aid, such as normal ear resonance and head diffraction, variations among ears, etc.

Chapter III contains a discussion of several experimental questions and the corresponding validity and reliability considerations appropriate to the present experimental design.

Chapter IV contains a description of a subjective-fidelity-rating experiment designed to assess the fidelity of three pairs of experimental high-fidelity hearing aids designed in accordance with the guidelines discussed in Chapter II.

Chapter V contains a summary of the fidelity-rating experiment, as well as a discussion of possible future research.

Several Appendices provide additional technical information.

I. SELECTIVE AMPLIFICATION AND SPEECH DISCRIMINATION

In 1940, Watson and Knudsen reported a series of experiments showing that a marked improvement in intelligibility could be obtained for both conductively and "perceptively" deafened individuals by the use of selective amplification. In contrast to simply mirroring the audiogram, they emphasized that to be successful, such amplification should "fit the frequency-intensity areas occupied by the amplified speech sounds into the diminished. . .auditory sensation area" (emphasis mine), and based their experimental procedures on measurements of equal loudness curves.

Unfortunately, later findings appeared to discredit the findings of Watson and Knudsen, and the selective amplification approach fell into disfavor. Both the Harvard Report (Davis et al., 1947) and the "MEDRESCO Report" in Britain (Radley et al., 1947) indicated that a single hearing aid design would produce about as good results as could be obtained on nearly all subjects. Indeed, one stated conclusion was that "The idea of individual selective amplification is fallacious" (Davis et al., 1946). The report of Shore, Bilger, and Hirsh

(1960) that it wasn't even possible to reliably distinguish among

hearing aids with different electroacoustical characteristics on the

basis of clinical measures (gain, discrimination in quiet and in noise)

added further weight to that conclusion, and the issue was considered

settled by many.

Not everyone accepted the findings of the Harvard study, of

course. Many clinics continued to use the Carhart method--which in-

cludes measurement of gain, discrimination in quiet, and discrimination

in noise--(Carhart, 1946, 1950) for selecting hearing aids, and most

hearing aid dealers knew from extensive experience that the electro-

acoustic characteristics of the hearing aid did make a difference to

many users.

Fletcher (1953) pointed out some of the problems in interpreting

the results of the Harvard study. One important problem was that the

real-ear or "orthotelephonic" frequency response of the master hearing

aid used was substantially different than the measured response re-

ported by Davis et al. (1947). Figure 1-1 shows the orthotelephonic

response--the ratio of aided to unaided eardrum pressure--which

Fletcher calculated for the "+6 dB/octave" setting of the master

hearing aid used in the Harvard study. Fletcher went on to demonstrate

that the findings of the Harvard study in individual cases would have

been predicted nicely on the basis of the Articulation Index (French

and Steinberg, 1949) calculations once the orthotelephonic response of

the master hearing aid was taken into account.

Only within the last decade, however, has substantial support for

selective amplification reappeared. Books by Victoreen (1960, 1973)
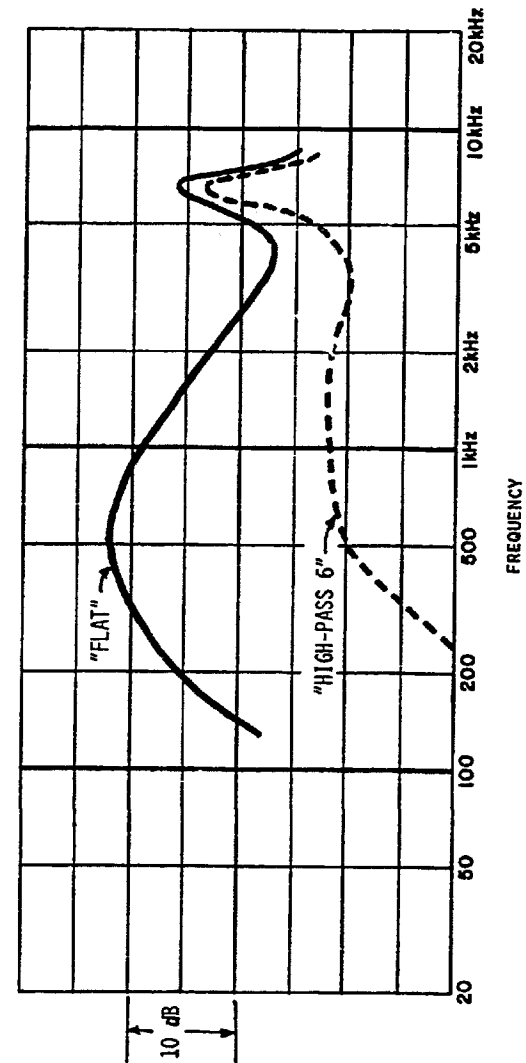
FIGURE 1-1 REAL-EAR (ORTHOTELEPHONIC) RESPONSE OF MASTER HEARING AID USED IN "HARVARD STUDY" (AFTER FLETCHER)

and Wallenfels (1967) described approaches for arriving at the proper frequency response. Following the reasoning of Watson and Knudsen (1940), Victoreen argued that the appropriate frequency response would place the amplified speech sounds at the most comfortable loudness level in each frequency band. Observing that the most comfortable loudness level generally lay roughly midway between threshold and discomfort for both normal and hearing impaired subjects, Wallenfels suggested that "bisecting the remaining hearing" was a simple way of choosing a frequency response which would place the amplified speech sounds near the most comfortable equal-loudness curve. In practice, both the Victoreen and Wallenfels methods will normally result in similar frequency response recommendations, as will the even simpler method proposed earlier by Lybarger (1944), in which the desired frequency response shape is obtained by plotting one-third to one-half the hearing loss (on a dB representation) as a function of frequency. (With the Lybarger method, a hearing loss which increased at the rate of 10 dB/octave with increasing frequency would require a hearing aid with a 3 to 5 dB/octave rising frequency response.)

Dodds and Harford (1968) reported that some sensorineurals with "ski-slope" high-frequency hearing loss could derive significant benefit from a "CROS" hearing aid which provided minimal gain at low frequencies and a sharply rising response above 1 kHz. Of the 20 patients tested with the open-mold CROS configuration, 6 demonstrated improvements ranging from 24 to 58 percentage units when their aided discrimination was compared to the "PBmax" score obtained with the relatively uniform response of the speech audiometer (the mean aided score for

those six was 79 percent, compared to the mean score of 46 percent obtained with the speech audiometer). Here was a clear demonstration that changing the frequency response of the reproduction system could make a dramatic difference in speech discrimination for some hearing-impaired subjects. The clinical findings of Dodds and Harford were subsequently corroborated in laboratory studies reported by Barfod (1972) and more recently by Skinner (1978).

A common defect in nearly all studies conducted before 1970 was the failure to obtain the real-ear frequency response of the hearing aid; the same shortcoming which Fletcher observed in the Harvard study. The difference between the response of a hearing aid as measured in a two cc coupler and the response of a hearing aid as experienced by a wearer had been discussed by Romanow (1942), LeBel (1944), and others in the case of body-worn hearing aids. Knowles (1959, 1967) described tentative correction curves for head-worn aids, based on the objective measurements made by Wiener and Ross (1946) of head diffraction and eardrum pressure in a progressive sound field. Not until the 1972 introduction of the KEMAR manikin (Knowles Electronics Manikin for Acoustic Research; Burkhard and Sachs, 1975), however, did it become commonplace for hearing aid research to take these corrections into account in the experimental design.

An exception was Fournier (1965), who reported the routine use of free-field Bekesy audiometry to measure the functional gain (the difference between aided and unaided thresholds) of hearing aids. Fournier and Rainville (1967) reported that the best speech discrimination normally resulted when the aided field audiogram was flat; i.e.,

when the real-ear response of the hearing aid mirrored the audiogram.

The first comprehensive study of selective amplification which was not marred by defects in experimental design—either in electroacoustic or psychoacoustic method—was only recently completed by Pascoe (1975). Pascoe used a binaural master hearing aid, incorporating wideband hearing aid microphones and receivers, along with a pair of one-third octave filter sets which made possible a wide variety of frequency responses. Perhaps most importantly, Pascoe measured the functional gain of the master aid for each of his subjects and included corrections in the filter settings so that the frequency responses he investigated represented the actual response experienced by the subject. The validity of these corrections was further tested by comparing the speech discrimination obtained by each subject in the "unaided" sound field condition[1] with that obtained with the "uniform amplification" settings of the master hearing aid: The differences were insignificant.

Pascoe's eight subjects had mild-to-moderate hearing loss, with an average audiometric slope of roughly +6 dB per octave (increasing loss at high frequency). The audiometric contour ranged from relatively flat to a maximum of +10 dB per octave slope. Only one subject had a hearing level greater than 60 dB at 4 kHz.

Using a counterbalanced presentation of word lists (each 50-word list was presented in 10-word segments separated by the time required

---

[1]The use of the term "unaided" to describe earphone or sound-field testing at presentation levels of 90 or 100 dB SPL is perhaps an unfortunate choice of terms. A better description might be "earphone aided" or "loudspeaker aided" testing.

to cycle through all the experimental conditions) in quiet and noise and with both a male and female talker, Pascoe found a consistent and statistically significant improvement in discrimination for all subjects when wideband selective amplification (mirroring the audiogram in this case) was compared to conventional hearing aid response (either the subject's aid or Pascoe's "aid simulation" condition): an average improvement of 18.4 percent over all talkers and conditions using high-frequency word lists, and a 20.9 percent improvement using PB lists in noise.

Part of the improvement Pascoe found was due to the greater effective bandwidth of the master hearing aid (.1 to 6.3 kHz) compared to the "aid simulation" condition, as evidenced by the average of ten percentage units improvement in discrimination scores obtained by his subjects with the "uniform amplification" condition compared to the aid-simulation condition. As might be expected, the subjects with a sloping loss did much better with the selective amplification than with uniform amplification; those with a relatively flat loss did about the same. In the latter case, of course, the frequency response of the "selective amplification" was little different from uniform amplification.

When the loss is severe at some frequencies and the audiogram slopes sharply, the available dynamic range is often too restricted to make full mirroring of the audiogram practical.

Skinner (1976) recently reported a comprehensive study of the effect of selective amplification on subjects with such a "ski slope" hearing loss. In an improvement on Pascoe's technique, Skinner

eliminated the use of hearing aids entirely by applying the frequency-selective filtering before the signal reached the loudspeakers. (The filtering included corrections for the frequency response of the loud-speakers themselves.) As a result, Skinner's findings are entirely free from questions regarding the true insertion gain and frequency response of the hearing aids and, thus, provide direct guidance for hearing aid design.[2] Skinner found that only 20 to 30 dB of high-frequency emphasis was practical in a hearing aid employing linear amplification, even for subjects with 50 to 60 dB greater loss at high frequencies than at low frequencies.

This result had been known qualitatively for some time, of course. As stated by Lybarger in 1944:

> I have found that full corrections are unsuitable in most (of those) cases. . .because amplification equal to the large hearing loss at some frequencies is so great that it causes annoyance and pain to the wearer of the hearing aid. Also, because of the small hearing range between the sound threshold and pain for nerve cases, a full correction that seemingly would bring the hearing loss curve back parallel with the normal threshold curve is too large, and a hearing aid having such a correction is very unpleasant for the user.

A way out of this dilemma would be some sort of electronic process-ing so that the subject could have a frequency response which provided both a mirroring of his threshold audiogram (for quiet sounds) and

---

[2]While Skinner's loudspeaker technique is undoubtably the tech-nique of choice for any basic-research study of the effect of frequency response on subject performance with linear amplification, it would be inappropriate for use in a study involving compression or Automatic Gain Control (AGC) unless the tests were carried out in a quiet ane-choic chamber: The effect of room noise and reverberation added after compression amplification may be substantially different than that of room noise or reverberation present in the signal before compression.

amplification to a comfortable loudness contour (a la Watson and Knudsen, 1940) for louder sounds. This is precisely what the multi-channel compression system discussed by Villchur (1973) can provide. In effect, Villchur's system attempts to mirror the hearing loss not only at threshold but also at the higher sensation levels where the recruitment which typically accompanies sensorineural hearing loss of cochlear origin acts to reduce the loss in loudness (Steinberg and Gardner, 1937). Villchur's system represents a combination of two types of signal processing acting simultaneously; one in the frequency domain and one in the amplitude domain.

Finally, extensive studies by Barfod (1976, 1978), and Lippman (1978) have shown that the appropriate choice of selective amplifica-tion can not only produce significant improvements in discrimination scores for subjects with moderate hearing impairments, but can equal the best multi-channel compression results when the test words are all presented at the same (optimum) level. In other words, when subjects have sufficient dynamic range to encompass the within-word variations in speech elements, the use of appropriate frequency-response shaping alone appears sufficient to optimize discrimination scores as long as the traditional constant-presentation-level speech tests are employed.[3]

In summary, it is now clear that some form of wideband selective amplification to compensate for the frequency dependence of a hearing

---

Although a discussion of experiments on compression amplification is beyond the scope of this chapter, it is perhaps worth noting Villchur's (1978) observation that traditional speech tests actually use effective-ly pre-compressed speech material, since the tests are recorded with the talker facing a vu meter and carefully monitoring his level to keep it constant.

loss is required to maximize speech discrimination. For subjects with mild or moderate hearing loss, a substantial dynamic range (defined here as the distance between threshold and discomfort or distortion levels) often remains, so even mirroring the audiogram is often practical. Whether such mirroring would be desirable or not is a question discussed below. [An expanded history of selective amplification was given recently by Lybarger (1978)].

II. OTHER FACTORS

The task of demonstrating a significant advantage for a given hearing aid is often more complicated than simply showing an improved speech discrimination for some subjects. Unfortunately, much more effort has been devoted to determining the influence of the electroacoustic characteristics of hearing aids on speech discrimination than to determining whether or not that was the most important question (or even if it was an important question) for the population under study. As Barfod (1972) so clearly states, a hearing aid is a personal prosthesis and, thus, the criterion for deciding whether one hearing aid characteristic is superior to another can only reasonably be chosen by the patient himself.

A. The Relative Importance of Speech Discrimination Testing

There are several indications that finding the electroacoustical characteristics of a hearing aid which maximize speech discrimination-- as commonly measured--is often not the most important consideration:

1. Generally speaking, hearing aid wearers don't voluntarily turn the gain up enough to maximize discrimination. The studies of Wallenfels (1967) and Martin (1973) of the gain settings used

in normal conversational settings suggest that the typical user with a 60 dB sensorineural hearing loss chooses to listen at a 20 dB Sensation Level (i.e., 20 dB above his threshold); a level well below the 30 to 40 dB required to maximize his speech discrimination (Tillman and Carhart, 1966).

2. The best estimates are that nearly 500,000 "MEDRESCO" body-worn hearing aids were consigned to dresser drawers because of their bulk (Knowles, 1968), even though they were designed to maximize speech discrimination in accordance with the findings of Medical Research Council report (Radley et al., 1946; the British equivalent of the "Harvard Report").

3. Watson and Knudsen (1940), Silverman and Taylor (1947), and Davis et al. (1947), all reported that new hearing aid wearers generally preferred the "smoother, more mellow" tone of a hearing aid with a limited frequency response and no high-frequency emphasis to the "harsh and unpleasant" sound of the amplification which yielded maximum discrimination scores. Old hearing aid wearers generally preferred the "more limited and inferior response of their own instruments" to the response of a wideband master hearing aid (Davis et al., 1947). Thompson and Lassman (1970) found approximately 80 percent of their subjects with sloping high-frequency sensorineural hearing loss preferred a "flat" frequency response even though 60 to 70 percent of them had better speech discrimination scores with high frequency emphasis. Essentially similar findings were obtained recently by Pascoe (1976, personal

communication): Most of his subjects preferred the uniform response setting of the CID master hearing aid, even though they obtained significantly higher speech discrimination scores with high-frequency emphasis.

B. Learning and Adaptation

The problem of demonstrating a meaningful "significant advantage" for a particular hearing aid characteristic is further complicated by habituation ("Wear it for a few weeks and you'll get used to it.") and learning effects. As anyone involved in sensory experiments is well aware, the central nervous system is capable of enormous adaptation and (most) subjects are capable of an incredible amount of learning. The time constant required for adaptation may range from seconds to months, depending on the complexity of the task, previous experience with similar tasks, and the condition of the central nervous system. Part of the "art" involved in selecting a hearing aid for a given individual is trying to make a good assessment of that individual's patience (Stutz, 1975). In many cases, the best thing to do is to provide the individual with a hearing aid that is less than "optimum" but which also provides a smaller departure from what the individual is accustomed to hearing. This can sometimes help avoid the "perfect" hearing aid sitting in the individual's dresser drawer because he won't put up with how it sounds for a long enough period to get used to it.

Some of the experiments with vision provide the easiest-to-understand illustrations. A fascinating review article appeared some years ago in Scientific American (Kohler, 1962) reporting the results

of several experiments relating to long-term visual adaptation. In one of these experiments, the subject wore glasses which caused everything to appear upside down. The glasses were taped to the subject's head so he had no other visual input for a period of weeks. At first, of course, he had difficulty walking, reading, writing, etc. After a period of some weeks, however, things had returned to something like normal for the subject, and he could read, walk, and even ride a bicycle without difficulty. After he had adapted to the inverted image, the glasses were removed. The subject then had to go through a several-day period of readaptation, during which time he had difficulty walking, couldn't ride a bicycle, etc.

Although one must always exercise great caution in applying any analogy between vision and hearing, several experiments indicate the auditory system works in a similar fashion:

1. The simplest experiment can be easily performed by anyone willing to wear a hearing aid for a few weeks. If the real-ear response of the hearing aid provides a large departure from the unaided sound quality, the initial reaction is usually one of "trying to hear speech in a sea of noise." The "coloration" of the sound provided by the limited bandwidth, response peaks, etc., in the frequency response of the hearing aid make it impossible to recognize and ignore even the most common background noises: those which we normally don't even notice unless they are forcibly brought to our attention (Rhodes, 1969). After a few weeks, however, someone with normal hearing finds the hearing aid(s) tends to subjectively disappear;

background noises recede again into the background, and
speech communication becomes easy again.

2. Anyone servicing hearing aids is familiar with the problems
which may result from restoring the frequency response of a
damaged hearing aid back to its original configuration. As
Watson and Tolan (1949) report: "Hearing aid service workers
have remarked that users can often detect a change in response
of 6 or 7 dB over half an octave and protest against their aid
'having been changed.'" Similar problems result from the use
of "loaner" hearing aids during an extended repair period
(Goldberg, 1965a). Even when the general frequency responses
of the original and loaner aid appear similar, it is not un-
common for a user to complain initially of difficulty in
understanding speech because the loaner is "noisy". . .and then
find after a two or three week period that the loaner seems
much better than his repaired original aid. (The writer has
had occasion to loan a hearing aid to a friend and hear for
himself the same comments.)

3. As discussed elsewhere, there is reason to believe that a
hearing aid whose frequency response exhibits sharp peaks
will be less useful than one without such peaks, but the
writer has had two experiences where smoothing the response
(by adding damping to the earmold tubing) for a long-time
hearing aid wearer resulted in a negative evaluation. The
ostensibly "improved" hearing aid "wasn't as good as before."

4. Lastly, the writer experienced a personal case of adaptation
strikingly similar to the upside-down visual-field adaptation.
After wearing an in-the-ear hearing aid with 8 kHz bandwidth
during the summer, he wore it almost every waking hour for
one week and, in a burst of enthusiasm for the project, wore
it to bed one night. Finally remembering to remove the hear-
ing aid before stepping into the running shower the next
morning, he was startled by an unusual auditory illusion. It
sounded as if someone had suddenly turned on an 8 to 16 kHz
octave-band noise source, located on the wall opposite the
shower head. With head movement, the source appeared to
localize at the (now unaided) ear. This illusion lasted for
several minutes before finally fading away, during which time
the lower-frequency noise from the shower remained normally
localized at its source. Considering the relatively short
period the writer had deprived that ear of (ultra) high-
frequency input, it is perhaps no surprise that new hearing
aid wearers sometimes require a period of weeks before
regaining anything like normal auditory spatial perception,
binaural squelch, etc.

C. Adaptation and Speech Discrimination

When the criterion for evaluating a hearing aid is the speech dis-
crimination score obtained with the aid, failure to allow for adaptation
and learning effects risks the possibility of sterile or erroneous con-
clusions. If the hearing aid does anything to provide correction for

high-frequency hearing loss, it may provide the subject with acoustical

cues which he has not heard for years. If the subject no longer has an

"auditory image" (Goldberg; 1960, 1965b) for those cues, this addition-

al information may come in as "noise" and actually decrease his dis-

crimination score in initial testing because of the distraction

produced. Indeed, Watson and Knudsen (1940) report a case in which uni-

form amplification was initially superior to selective amplification

(high-frequency boost) but after a few weeks exactly the opposite was

true. A similar case was seen in the files of one of the Chicago-area

hearing clinics. An octogenarian with bilateral mild-to-moderate

sloping hearing loss and no previous hearing aid experience was tested

with several high-frequency emphasis hearing aids. Her initial score on

the first aided test was 16 percent--well below the 36 percent obtained

with the more or less uniform amplification provided by the speech audio-

meter. After testing with several other high-frequency emphasis aids,

her retest score with the initial aid had climbed to 32 percent. After

purchasing the same model as the initial aid, she was seen for retest

five weeks later and obtained a score of 52 percent in two successive

tests. Because of the lack of careful controls, such a case history

must be viewed with caution, but it is consistent with a fair amount of

anecdotal evidence.

That the human auditory system can adapt to a modified set of

speech cues, however, is illustrated by the Licklider and Pollack

(1948) findings with speech which had been subjected to infinite peak

clipping. Initial discrimination scores for their listeners averaged

71 percent, increasing gradually over the 25 practice sessions to a

value of 95 percent. Further tests with novel word lists led the in-

vestigators to conclude that 15 percentage units of the improvement

were a result of the listeners having learned to identify words cor-

rectly despite severe distortion. A hearing loss represents a differ-

ent type of distortion than infinite peak clipping, but learning to

extract meaning from previously unheard (or forgotten) high-frequency

speech cues may require a similar amount of practice time.

D. Effort and Fatigue

In some cases, factors which may be enormously important to the

hearing aid user are difficult or impractical to measure. One of these

is what Watson (1944) referred to as: "The Neglected Factor of Effort

and Fatigue in Hearing." As Watson observed, "An excellent speech

intelligibility test. . .measures only whether the subject understands

or not without measuring how easily he understands." He went on to

suggest a several-hour test (admittedly impractical) which would allow

the evaluation of the effort and fatigue factors. Watson's analogies

are difficult to improve upon:

> A factor which renders any brief list of words or sentences
> misleading as a test of hearing is the ability of the ear to
> accommodate for wide deficiencies in loudness, limited frequency
> range, distortion or other defects for short periods of time.
> The eye with a certain amount of added effort can read a printed
> page under substandard illumination. Yet no one would say that
> the illumination was satisfactory or visual acuity normal merely
> because the subject had succeeded in reading a page of text
> correctly. Similarly, through a distorted or misfitted lens one
> can by increased effort read a page of text or name a series of
> objects correctly. Yet no one would pretend that such a lens was
> acceptable for that reason.

> Persons with normal hearing can generally understand speech in
> an auditorium with poor acoustics even though they are annoyed
> by reverberation. They would generally score well in a brief

intelligibility test, though they would have to make an effort much greater than that necessary to understand the same words under favorable acoustic conditions. Acoustical engineers are well aware that the constant effort necessary to understand speech under poor acoustic conditions leads to fatigue. They would not insist that the acoustics of an auditorium were acceptable merely because a number of subjects made an excellent score during an hour's speech intelligibility test. They would proceed, instead, to make careful acoustic measurements independent of human speech and the accommodation powers of the ear. The basic data obtained would be used to formulate a plan for correcting acoustic conditions in the auditorium.

Certainly one factor in determining the fatigue accompanying the use of a hearing aid is the maximum output of the hearing aid relative to the discomfort (or pain!) level of the user. Many experienced hearing aid dealers consider this the most important factor in determining whether or not a given individual will be a successful hearing aid wearer (Wallenfels, 1967). Briskey (1976) asserts that finding the correct SSPL (Saturation Sound Pressure Level) represents 70 percent of the successful hearing aid fitting. (The "overpowering" of the user by University Clinic's hearing aid selection processes was for yea- a favorite joke among successful hearing aid dealers across the country.)

Another factor is likely to be a frequency response containing sharp peaks. Although little concrete evidence exists beyond the old Bell Lab's findings that a resonant peak in a transmission system could reduce speech discrimination (articulation) (Fletcher, 1953), the deterioration in sound quality brought about by multiple peaks is readily apparent to any careful listener. With the reduced dynamic range accompanying most hearing impairments, the presence of peaks produces a more direct liability in that they limit the amount of gain

which can be employed before occasional vowel formant peaks or other sounds (at the frequency of the peak) are made uncomfortably loud. Similarly, listening to a hearing aid with pronounced 15 to 20 dB high peaks has been likened to listening to an electronic expander (which exaggerates the normal differences between loud and soft sounds). As discussed below, this is precisely opposite to the characteristic which may be needed in a hearing aid. Indeed, the presence of sharp peaks in the frequency response may mitigate the effectiveness of compression amplification!

E. Cosmetics and Wearability

The relative importance of such non-acoustical factors as size, weight, and appearance has been measured in the marketplace. Although at least one writer has suggested that a 15 pound hearing aid would be acceptable if only it had adequate sound quality (Rosenthal, 1976), experience indicates that such a hearing aid would be totally unacceptable to the vast majority of users regardless of its sound quality. Body-worn hearing aids with wide bandwidth and good sound quality have been available for a quarter century, yet even when only narrowband frequency responses were available in headworn aids, over 80 percent of all hearing aids sold in the United States were of the (generally more expensive) headworn variety (Lybarger, 1974). Even when body-worn hearing aids have been provided free of charge, as in the case of the MEDRESCO aids discussed earlier, the majority of potential users have chosen to do without any hearing aid or pay for a headworn aid out of their own pocket.

F. Cost

However one goes about obtaining a hearing aid in today's market, the total cost of a single aid will be comparable to that of a major appliance. Similarly, not only the initial purchase price but the operating cost (determined by the battery drain) and maintenance cost are all important factors in the total cost. A discussion of the price-demand curve is beyond the scope of this paper, but no list of the factors on which people judge the utility of hearing aids would be complete without inclusion of the cost factor.

G. Sound Quality

For the hearing aid user whose hearing loss is so severe that he has difficulty understanding speech even with a hearing aid, it is reasonable to assume that the sound quality of the hearing aid is a secondary consideration. As discussed by Killion and Carlson (1970), however, there exists a large number of hearing aid users with mild-to-moderate hearing loss for whom speech discrimination is no problem provided the speech sounds are made loud enough. For them, the other factors discussed above are more important than any attempt to further improve speech discrimination with the hearing aid. Although there is substantial market evidence that the sound quality per se is less important than the cosmetic and wearability factors, there is also a fair amount of anecdotal evidence that there are many borderline hearing aid (non-)wearers who would no longer put up with the frequent inconveniences of their hearing loss if a truly high-fidelity head worn hearing aid were available to them at a reasonable price.

III. RATIONALE FOR A HIGH-FIDELITY HEARING AID

Fortunately, there is reason to believe that many of the traditional difficulties in evaluating hearing aids were caused by technical limitations in the hearing aids themselves. The author's listening experience with hearing aids having a wideband and smooth real-ear response (as defined below), for example, is that the adaptation time required before background noises cease to be troublesome is reduced from the days or weeks required with many conventional hearing aids to essentially zero: Sounds localize normally and the aid tends to subjectively disappear immediately unless one consciously listens for the difference in sound quality. (This has been true even with 10 to 15 dB of high-frequency emphasis included in the frequency response.)

Similarly, part of the reason for the common finding that a uniform frequency response is preferred to one with high-frequency emphasis may relate more to the overload characteristics of the hearing aid than to the frequency response per se. Although the majority of hearing aids now employ some form of low-distortion automatic volume control (AVC) or compression limiting to limit the maximum output before voltage clipping occurs, this was not true as recently as a decade ago. When voltage clipping alone limits the maximum output, it is likely to occur first at the higher frequencies because the electrical impedance of the earphones (receivers) used in hearing aids rises with frequency. Adding a substantial high-frequency emphasis to the hearing aid may increase intelligibility, but it also increases the likelihood that clipping will occur (or user discomfort levels will be

exceeded) during normal conversations. The resulting distortion can produce a raucous and unpleasant sound; something like trying to listen through the sound of a house burning down next door. Peterson (1951) studied this phenomenon in one hearing aid, whose treble-boost tone control produced a better looking frequency-response curve but a poorer user rating. (The alternate approach--increase the maximum sound pressure output capability of the hearing aid at high frequencies by choosing a lower earphone impedance--is often even less acceptable because of the limited dynamic range exhibited by precisely the individual who needs the high-frequency emphasis. An increased maximum output would often exceed his threshold of discomfort at high frequencies and/or subject him to ear distortion.) Indeed, Johansson, Sjogren, and Nyman (1974), in a study of the effect of amplitude distortion in hearing aids on the intelligibility of speech, concluded that: "The perceived sound quality deteriorates much faster than the speech intelligibility."

Support for the hypothesis that it is not the high-frequency emphasis per se which subjects dislike came from the study of Villchur (1973), who used substantial high-frequency emphasis coupled with a low-distortion, two-channel compression system to reduce the dynamic range of (especially) the high-frequency elements in speech. Not only did his subjects obtain significantly higher speech discrimination scores with that system, but they universally _preferred_ it to the hearing aids they were wearing. The 1975 Pascoe study mentioned earlier, incidentally, employed a Master Hearing Aid whose 120 dB SPL maximum output was determined by voltage clipping. The preference of his

subjects for the uniform amplification over the high-frequency-emphasis amplification, therefore, is consistent with the current hypothesis.)

The above arguments suggest that a high-fidelity hearing aid would be advantageous for a significant percentage of hearing aid candidates. Various authors have made similar suggestions in the past: Corliss, Kobal, and Berghorn (1960), Miller and Niemoeller (1967), Tillman, Carhart, and Olsen (1970), and Harris (1971). Corliss et al., Miller and Niemoeller, and Tillman et al. based their recommendation on their findings that in the presence of competing noise, a deterioration in intelligibility was produced by the hearing aids they tested when compared to a high-fidelity system. Harris (1971) based his plea on the observation that many desireable sounds other than speech (fifes, surf, birds, et.) lie outside the passband of most hearing aids. (In the sense used by these authors, a high-fidelity hearing aid would be a high-fidelity sound reproduction system with added gain. As discussed in Chapter II, below, a more sophisticated design may be desireable.)

Recent clinical evidence suggests hearing aids have improved sufficiently since the time of the (body-worn) hearing aids tested by Corliss et al., Miller and Niemoeller, and Tillman et al., that those early findings may no longer be applicable. Katz and Longinotti (1977) found that 20 percent of subjects tested for hearing aid selection in the Northwestern University Hearing Clinic obtained a significantly higher discrimination score with the tested hearing aid than obtained under earphones with a "high-fidelity" speech audiometer; only nine percent obtained a significantly poorer score.

Even if the majority of users obtained no better speech discrim-
ination scores with a high-fidelity aid, however, the other factors
discussed above argue for the utility of such an aid if it can be made
small enough to be cosmetically acceptable. In addition, such an aid
would provide a more convenient starting point than is currently avail-
able from which to make modifications in electroacoustical character-
istics. Perhaps the greatest benefit of a demonstration that a high-
fidelity hearing aid was practical would be the encouragement it might
provide to research aimed at discovering what a hearing aid should do
("if technology were only available"), rather than research aimed at
illustrating the defects in previous hearing aid designs or finding
better methods for selecting among existing hearing aids.

The "problem" as the author sees it is (a) no high fidelity hear-
ing aid is currently available, and (b) no one knows for sure whether
or not users would benefit sufficiently to make the extra effort of
building high-fidelity aids worthwhile. (Since the time this chapter
was originally written, at least one commercial high-fidelity hearing
aid has been made available. It is commented upon in Chapter V.)

All of these considerations lead to the following hypotheses:

1. It is now possible to design a small head-worn hearing aid
    having a smooth, wideband frequency response whose sound
    quality will be judged "high-fidelity" by normal listeners.

2. Such a hearing aid will be judged preferable to their own
    (conventional) hearing aids by the majority of users with
    mild-to-moderate hearing loss.

3. Subjects with mild-to-moderate downward sloping hearing loss
    will prefer an aid with high-frequency emphasis (over one
    with uniform response), provided it has a smooth frequency
    response and low distortion at high frequencies.

Only the first hypothesis can be adequately tested in a laboratory ex-
periment. The second two hypotheses can be adequately tested only with
extended use of such head-worn hearing aids by users with hearing loss.
A careful test of the first hypothesis comprises the essence of this
dissertation.

CHAPTER II

DESIGN CONSIDERATIONS FOR HIGH FIDELITY HEARING AIDS

Many sound-reproduction system characteristics are considered most naturally in the frequency domain, while others are considered most easily in the amplitude or time domain. That organization is used here. Considerations of bandwidth, frequency response smoothness, external-ear resonances, etc., are treated under the general heading "Frequency Response Requirements". Considerations involving peak undistorted output, non-linear distortion, automatic-gain-control input-output characteristics, and level-dependent frequency-response modifications are treated under the heading "Amplitude Response Requirements". Finally, the "transient response" required of a high-fidelity hearing aid is treated under the heading "Time Response Requirements."

Generally speaking, the certainty with which each requirement can be determined on an a priori basis decreases steadily as one moves from the frequency to the amplitude to the time domain. In all cases, however, the requirements imposed on a high-fidelity hearing aid can be met using existing technology. Where possible, an example of a transducer and/or amplifier configuration designed to meet each requirement is given as that requirement is discussed.

I. DEFINING THE GOAL

The term "high fidelity" is more easily defined for someone with normal hearing than for someone with a hearing impairment. In the latter case, we will accept Barfod's (1972) definition: "One could say that the ideal hearing prosthesis was an instrument which gave the wearer the same perception of external stimuli as a normal hearing person would have".

Although that goal may often be physically unrealizable, it can provide the proper outlook as we view (1) the requirements for a high fidelity hearing aid for someone with normal hearing, (2) the additional requirements imposed by hypothesizing even a simple form of sensorineural hearing loss, and (3) the overall goal resulting from the combination of (1) and (2).

A. High Fidelity "Hearing Aid" for the Normal Hearing

As a first step, it is useful to consider the requirements imposed on a hearing aid if it is to achieve high fidelity as a unity-gain sound-reproduction system. An ideal high-fidelity system would be acoustically--and more importantly, subjectively--transparent. That is, the listener should receive the same auditory sensation with the high-fidelity system interposed between him and the original source of sound as he would have received listening directly.

Within the accuracy required by the limits of auditory discrimination, a high fidelity system must thus deliver the same pressure at the listener's eardrum as he would have received without the system interposed. /The above statement ignores tactile sensations and assumes that the stimulus delivered to the auditory system can be defined solely in terms of eardrum pressure. The latter assumption is both intuitively satisfying and consistent with the majority of recent

experimental evidence summarized by Killion (1978a)7.

Interestingly enough, a hearing aid stands a better chance of meeting this requirement than other types of sound-reproduction systems. The acoustics of the playback room unavoidably alter the sound in loudspeaker reproduction. The absence of head-movement-derived localization cues tends to limit the naturalness of headphone listening (even with good binaural recordings). In contrast, the headworn hearing aid can be worn in a live situation, thus avoiding the second-room-acoustics problem while providing normal head-movement-derived cues.

B. High Fidelity for the Hearing Impaired

An individual with no useful high-frequency hearing would presumably derive little benefit from wide-frequency-range sound reproduction, and so it seems reasonable to restrict our considerations to those with no more than a mild-to-moderate hearing loss at any frequency. Because appropriate surgery is now available to most conductive-loss sufferers, we will further restrict our attention to the so-called sensorineural hearing loss of cochlear origin (cochlear impairment). Finally, we will assume a particular form of cochlear impairment; one characterized by a loss of sensitivity for low-level sounds coupled with essentially normal hearing for high-level sounds. This last restriction is a strong one and requires further comment.

It has been known for some time that some individuals with mild-to-moderate cochlear impairment (once called "nerve loss") have excellent speech discrimination and excellent sound-quality judgment

capabilities for sounds which were sufficiently intense. Nixon (1945) reported that some of the individuals found most competent in exercising judgment of (radio broadcast) program quality at NBC had hearing impairments as great as 40 dB at 4000 Hz and above. More recently, Punch (1978) found no statistical difference between hearing-aid sound-quality judgments obtained from ten subjects with normal hearing and ten subjects with sensorineural hearing loss, even though some subjects had moderate-to-severe losses at high frequencies.

Such users were discussed by Steinberg and Gardner (1937) who observed that many of those with impaired hearing had essentially normal hearing for high-intensity sounds, but a loss of normal sensitivity for low intensity sounds. This phenomenon is commonly referred to as loudness recruitment. Figure 2-1 (from Scharf, 1978) illustrates this phenomenon for a "typical listener with a 40 dB hearing loss due to a cochlear impairment". The findings of nearly complete recruitment (normal loudness sensation for high-level sound) in subjects with cochlear impairment is so universal, in fact, that some form of loudness-recruitment test is routinely included in diagnostic test batteries aimed at excluding the possibility of hearing loss caused by a tumor along the eighth cranial (auditory) nerve (Hood, 1977). A substantial amount of evidence has accumulated since the Steinberg and Gardner study to indicate that many (but not all) individuals with mild-to-moderate cochlear impairments may have essentially normal high-level hearing. In addition to pure-tone loudness perception, the following attributes of hearing have been found to be within normal limits at sufficiently high intensity levels, even in the presence of
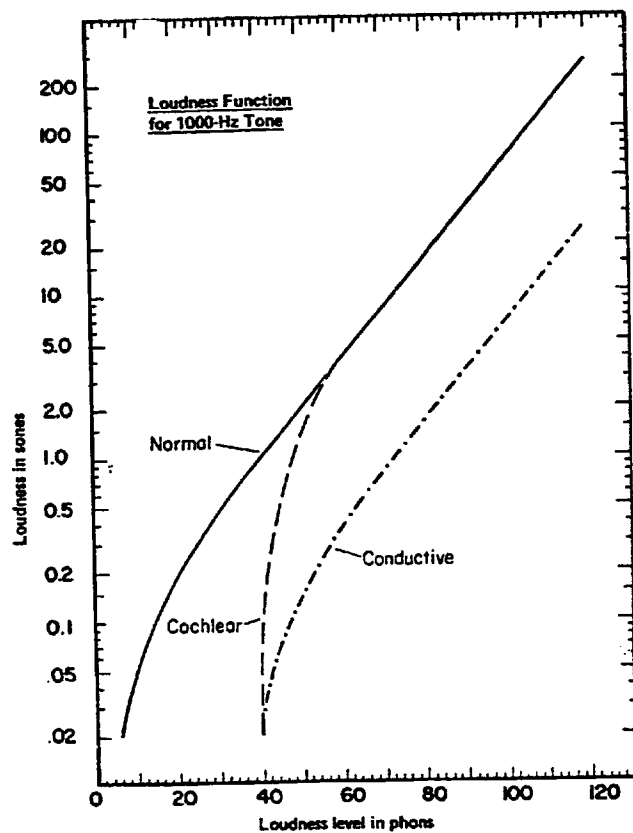
Figure 2-1   Loudness of a 1000-Hz tone as a function of loudness level for the normal listener, the typical listener with a cochlear impairment (and recruitment), and the typical listener with a conductive impairment.

(From Scharf, 1978.   Reproduced by permission of the Author and the Danavox Jubilee Foundation)

mild-to-moderate cochlear impairment:

1.   Frequency selectivity as determined from the "Fletcher Critical Band" which can be inferred from tone-in-noise masking experiments (Palva, Goodman and Hirsh, 1953; Jerger, Tillman and Peterson, 1960).

2.   Frequency selectivity as determined from psycho-physical tuning curves (McGee, 1978a).

3.   Frequency discrimination (dlF) (Gengel, 1973).

4.   Loudness summation for complex sounds (with cochlear impairment below 50 dB) (Scharf and Hellman, 1966).

5.   Loudness discrimination (dlI) (Scharf, 1978).

6.   Localization abilities (see the summary given by Scharf, 1978).

In addition to the psycho-physical data, recent evidence indicates that the high-level electrical potentials in the cochlea may be normal in some cases of mild cochlear impairment. The whole-nerve action potential recorded in laboratory animals (Wang and Dallos, 1972) or recorded from the ear canal in humans (Berlin and Gondra, 1976) is often normal. Indeed, both the wave form and latency of the entire Brain-Stem Evoked Response may appear entirely normal at high levels in some individuals (McGee, 1978b). Similarly, data obtained on laboratory animals with drug-induced outer hair cell damage (Dallos et al., 1977) indicate that it is possible to have normal bandwidth for both the psycho-physical tuning curve and the classical single-unit tuning curve obtained from fibers of the auditory nerve.

Several caveats are in order. First, and most important, there appear to be many types of cochlear impairment (see, for example, Schuknecht, 1974), some of which result in highly abnormal speech discrimination, psycho-physical tuning curves, etc. Such abnormalities are dramatically illustrated in the physiological data of Evans (1978). Moreover, even where a given individual appears to have normal high-level hearing by one measure, he may well have other, unmeasured, high-level abnormalities.

Even so, it seems reasonable to hypothesize the existence of individuals whose sensorineural hearing impairment is restricted to a loss of sensitivity for low-level sounds, and whose high-level hearing is essentially normal.

C. The Overall Goal Summarized

Assume, therefore, that we wish to design a high fidelity hearing aid for our hypothesized individuals. The appropriate high-level characteristics are self evident: the hearing aid should do absolutely nothing. Stated differently, the hearing aid should be "acoustically transparent"--in effect, subjectively disappear--for high-level sounds. (This conclusion is based on the assumption that an individual with completely normal hearing would not benefit from a hearing aid in most situations.)

Given these assumptions, the ideal hearing aid characteristic is seen to be that of a unity-gain high-fidelity sound reproduction system, as defined above, for high-level sounds.

For low-level sounds, on the other hand, the hearing aids must as a minimum provide sufficient amplification to make quiet sounds audible. If the hearing loss for low-level sounds is frequency dependent, as in the case of what is commonly called a "sloping hearing loss", more low-level amplification may be required at some frequencies than at others, implying a level-dependent frequency response may be required (Goldberg, 1960, 1972; Villchur, 1973; Barfod, 1976).

Lastly, the automatic-gain-control circuitry required to provide the variable-gain and variable-frequency-response amplification dictated by the requirements of the last paragraph should operate as unobtrusively as possible in order to provide high-quality sound reproduction at all sound levels.

In the next three sections, we will consider the specific Frequency-Response, Amplitude-Response, and Time-Response performance required of a high fidelity hearing aid.

II. FREQUENCY RESPONSE REQUIREMENTS

As might be expected, the sound pressure available to the microphone of a hearing aid is not the same as the pressure normally generated at the eardrum by a sound source. As demonstrated by Wiener and Ross (1946), head diffraction and external-ear ("earcanal") resonances combine to produce substantially greater eardrum pressure levels than those present in an oncoming sound field. At the roughly 2.7 kHz resonance of the outer ear, this gain amounts to 15 to 20 dB. The pressure at the microphone inlet of a headworn hearing aid, on the other hand, will generally be only about 5 dB greater than that in the

sound field, depending somewhat on the exact location of the micro-
phone (Madaffari, 1974; Kuhn and Burnett, 1977). Thus, the hearing
aid must provide some 10 to 15 dB of acoustic gain at 2.7 kHz in order
to compensate for the loss of gain the external ear itself provides
before it is blocked by the plastic earmold. Little compensation is
required at low frequencies, on the other hand, where the eardrum
pressure and the pressure available to the hearing aid microphone are
essentially the same.

It is useful at this point to formally introduce the term _inser-
tion gain_, which is the ratio of eardrum pressure produced by a hearing
aid to the eardrum pressure produced without the hearing aid (Dalsgaard
and Jensen, 1974). Expressed in dB, the insertion gain of a hearing aid
is the difference between aided and unaided eardrum sound pressure
levels. /Similar terms are orthotelephonic gain, etymotic gain and
functional gain. The term functional gain has generally been reserved
for subjective measurements of insertion gain, such as the aided-
unaided threshold-difference method (Pascoe, 1975), whereas the others
generally imply physical measurements. See Burkhard (1978) for further
discussion of these terms./

Recall that at high levels we wish the hearing aid to deliver the
same pressure at the listener's eardrum--at all frequencies and for all
angles of incidence of the original sound--as he would have received
without the hearing aid interposed. Stated in terms of insertion gain,
the insertion-gain frequency-response curve should be flat at high
levels. The frequency response of the hearing aid itself must be
tailored to produce this result.

A. Coupler Response for Flat Insertion Gain

For engineering purposes, the required frequency response tailor-
ing is best defined in terms of the coupler response of the hearing aid.
In addition to the factors discussed above (loss of external-ear reso-
nance and hearing aid microphone location), the COupler Response for a
Flat Insertion Gain (CORFIG) curve will be influenced by the choice of
couplers, the selection of reference sound-field conditions, and dif-
ferences among individuals in terms of their external ear resonances,
eardrum impedances, etc.

A great deal of simplification is provided if we ignore individual
differences and choose a manikin of average anthropometric dimensions
and a coupler ("occluded-ear simulator") which approximates the acoustic
impedance of an average ear. The KEMAR manikin (Burkhard and Sachs,
1975) meets the first requirement, while the modified Zwislocki (1970)
coupler meets the second requirement up to 7 or 8 kHz (Sachs and
Burkhard, 1972).

Figure 2-2 shows the Zwislocki-coupler CORFIG curves for these
microphone locations and three sound-field reference conditions, based
on measurements with a KEMAR manikin (Killion and Monser, 1979). The
inset drawings illustrate the three different locations for the micro-
phone inlet: Over-The-Ear, In-The-Ear, and In-The Concha. The Over-
The-Ear and In-The-Ear locations are similar to those found in headworn
hearing aids. (The microphone inlet in an In-The-Ear aid is approxi-
mately flush with the plane of the pinna since the concha is generally
filled by the aid itself.) The In-The-Concha microphone location is
similar to that described by Berland and Nielsen (1969), in which the
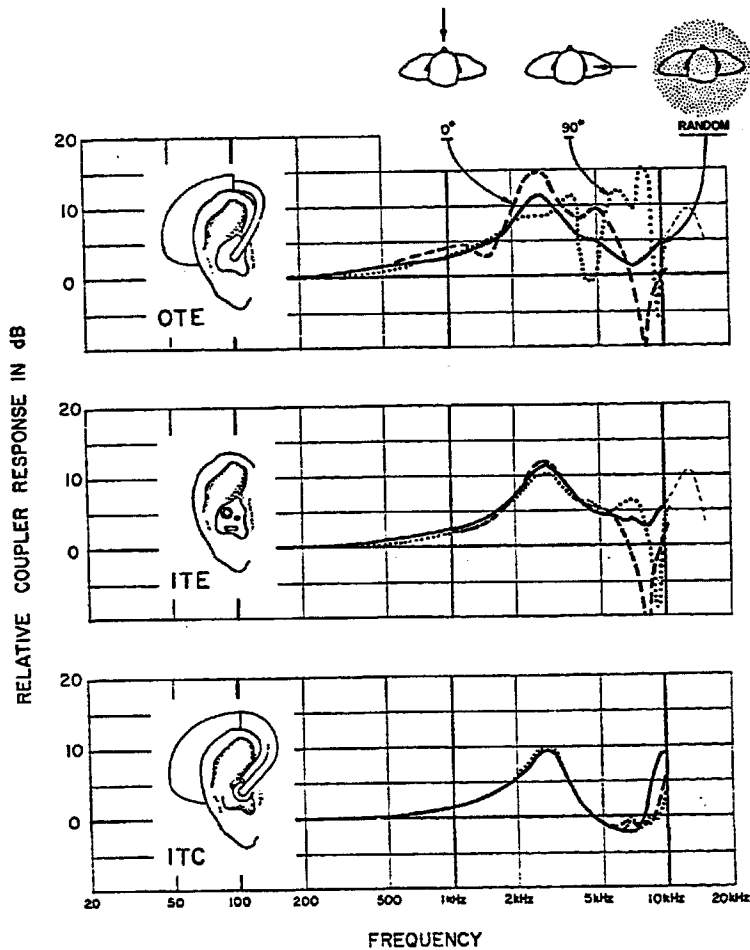
FIGURE 2-2 REQUIRED COUPLER RESPONSE FOR FLAT INSERTION GAIN

39

microphone inlet of an Over-The Ear hearing aid was extended down into the unoccluded concha in front of a "phantom" or "canal-lock" earmold.

The curves shown in Figure 2-2 agree (after inversion) quite well at comparable angles of incidence with the curves shown by Burkhard (1978) for angles of $0°$, $90°$, $180°$ and $270°$. The principle feature in all curves is the maximum response required at the roughly 2.7 kHz resonance frequency of the unoccluded external ear.

1. The role of the pinna and concha

As illustrated in Figure 2-2, the exact shape of the CORFIG curve generally depends on the direction from which the sound is coming. The effect of the pinna flange, for example, is to increase eardrum pressure by a few dB at high frequencies for sounds arriving from the front, and to reduce substantially the eardrum pressure for high-frequency sounds arriving from the rear "due to interference between the direct wave and a scattered wave from the edge of the pinna flange" (Shaw, 1974). For microphone locations outside the pinna, such as the Over-The-Ear, forward-looking-inlet location discussed above, very little of this directional dependence is contained in the sound available to the microphone. This loss was graphically illustrated in the data of Berland and Nielsen (1969), who compared the sound pressure available to microphones located behind, over, and in the ear for sound fields at six angles of incidence.

Even when the microphone is located in an In-The-Ear hearing aid, the presence of the earmold filling the concha substantially reduces the effectiveness of the pinna. Only when the microphone inlet is located directly in front of the blocked ear canal entrance are the directional effects preserved in their entirety. In the latter case, Shaw (1974)

indicates that the directional dependence of the differences between eardrum pressure levels and "blocked meatus" pressure levels in a free field is less than 1 dB up to 9 kHz and less than 5 dB up to 12 kHz. The concha itself can play an important perceptual role above 4 kHz for sounds arriving in the vertical (midsaggital) plane, i.e., from above or below the horizontal (Butler and Belendiuk, 1977).

With regard to auditory localization, it is clear than any of the microphone locations commonly used in binaural hearing aids will pre-serve the basic interaural time and intensity difference so important to binaural localization (Licklider, 1949), and the changes in inter-aural phase and intensity caused by head motion which are so important to the externalization of the sound (Wallach, 1940). What is not so clear is the relative importance of the pinna and concha to everyday localization, auditory spatial perception, and the binaural squelching of noise and reverberation.

Informal experiments conducted by the writer and others to assess the importance of these factors have included taping the pinna tightly to the side of the head, extending the earcanal with bent plastic tubes to provide a sound inlet at the Over-The-Ear location, and extended listening tests with actual hearing aids. These indicate that the pinna and concha provide relatively weak cues compared to those pro-vided by head movement and interaural time and intensity differences. Such a conclusion is consistent with anecdotal evidence indicating that many individuals with artificial pinnae wear them only on social oc-casions, i.e., for cosmetic reasons.

2. Reference sound-field condition

With regard to the frequency response, it is clear from Figure 2-2 that it is impossible to design a conventional In-The-Ear or Over-The-Ear aid which will have a perfectly flat frequency response for the user (i.e., an insertion gain which is frequency independent) at all angles of incidence. Where speech discrimination in face-to-face (near-field condition) listening situations is the dominant considera-tion, the appropriate reference condition is probably a 0°-incidence sound wave.

Where sound quality is the dominant criterion, other considera-tions apply. Most home and concert listening is done where the reflec-ted energy substantially exceeds the energy arriving directly from the sound source (Olson, 1967). Under those circumstances, the appropriate design compromise for a high fidelity hearing aid would appear to be a flat insertion gain for random-incidence sound. This conclusion is consistent with the results of the psychoacoustic experiments reported by Schulein (1975).

B. Bandwidth Requirements

After considering the available information on a) hearing sensi-tivity and frequency limits of hearing of typical listeners, b) mea-surements of the discomfort level of sound, c) measurements of room noise in a wide variety of locations, and d) measurements of the fre-quency limits in the maximum and minimum levels of speech, orchestral music and various instruments of the orchestra, Fletcher (1942) con-cluded that "substantially complete fidelity in the transmission of

orchestral music is obtained by use of a system having a volume range

of 65 dB and a frequency range from 60 to 8000 cycles per second".

Olson (1957) concluded that "the reproduction of orchestral music with

perfect (emphasis mine) fidelity requires a frequency range of from

40 to 14,000 cycles and a volume range of 70 dB". Both judgments were

based on similar data, primarily that of Snow (1931), and the differ-

ences reflected the nearness to perfection required. Snow's listeners

gave a judged quality rating slightly in excess of 90 percent to a sys-

tem whose frequency range extended from 60 Hz to 8000 Hz. With a 40 Hz

to 14,000 Hz range, the judged quality was approximately 100 percent

compared with unrestricted reproduction.

The results of Gannett and Kerney (1944) indicate that restriction

of the upper frequency range from 15,000 down to 8000 Hz represents

only two difference limens. Half of their observers could detect the

restriction from 15,000 to 11,000 Hz (one limen) and again from 11,000

to 8000 Hz. A frequency response extending from 60 Hz to 8000 Hz thus

appears a reasonable goal for a high-fidelity hearing aid.

An 8-kHz bandwidth has been within the capabilities of wideband

subminiature microphones for nearly a decade (Killion and Carlson,

1970). Figure 2-3, for example, shows the wide variety of frequency

response curves currently available in subminiature microphones; in-

cluding response curves which extend to 16 kHz and beyond.

An 8-kHz bandwidth is well within the capabilities of subminiature

wideband earphone such as the Knowles BP-series receiver (Carlson,

Mostardo and Diblick, 1976), but the choice of the tubing and earmold

required to couple the earphone to the ear is critical if this bandwidth
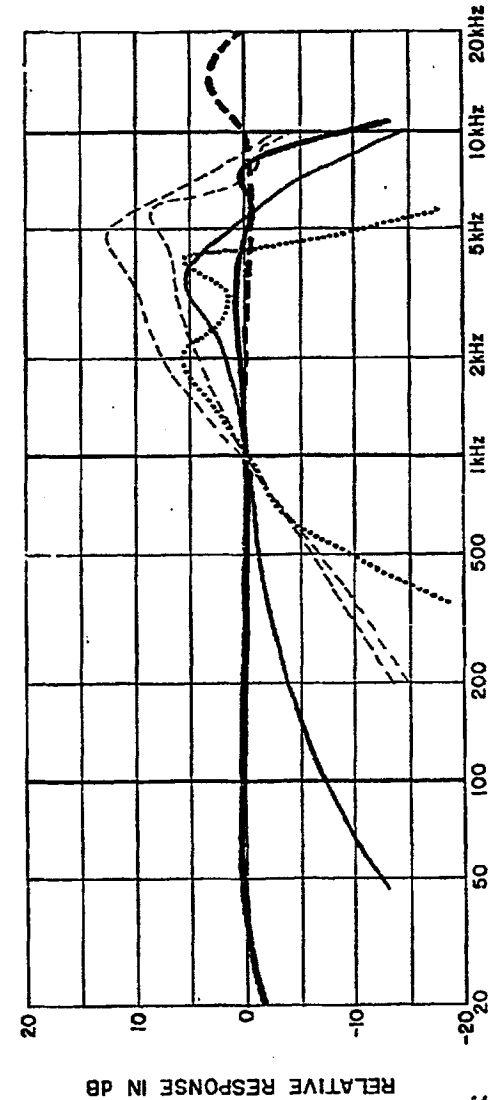
FIGURE 2-3 SOME RESPONSE CURVES AVAILABLE IN .01 CUBIC INCHES OR SMALLER MICROPHONES

RELATIVE RESPONSE IN dB

is to be realized in practice (Knowles and Killion, 1978). Figure 2-4 shows a coupling system as employed in an Over-The-Ear hearing aid designed for extended high-frequency response, but used with a conventional earmold.

1. The "8CR" earmold

The acoustic impedance of the subminiature earphone is nearly 100 times greater than that of the load presented by the earcanal and eardrum impedance. The use of increasing diameters in the coupling system as the earmold tip is approached provides an impedance transformation which can be exploited to improve the high-frequency response of the earphone and to simplify the task of smoothing the response peaks introduced by the presence of standing waves in the coupling system (Lybarger, 1972; Knowles and Killion, 1978; Killion, 1979b).

With the appropriate combination of coupling system and acoustic damping, it is possible to achieve an earphone-earmold response which provides both an 8-kHz bandwidth and a maximum response at 2.7 kHz. One such combination is illustrated in Figure 2-5 which shows the frequency response of a BP-series earphone coupled to a Zwislocki coupler through the Over-The-Ear mounting shown in Figure 2-4 combined with an "8CR" ("8 kHz; Canal Resonance") earmold. The construction of the 8CR earmold is shown as an inset in Figure 2-5.

2. The "16KM" earmold

Nearly two years ago, one of the writer's colleagues (E.L. Monser) developed a damping coupling system suitable for In-The-Ear hearing aids, which provided a 16-kHz bandwidth for the earphone-earmold combination.
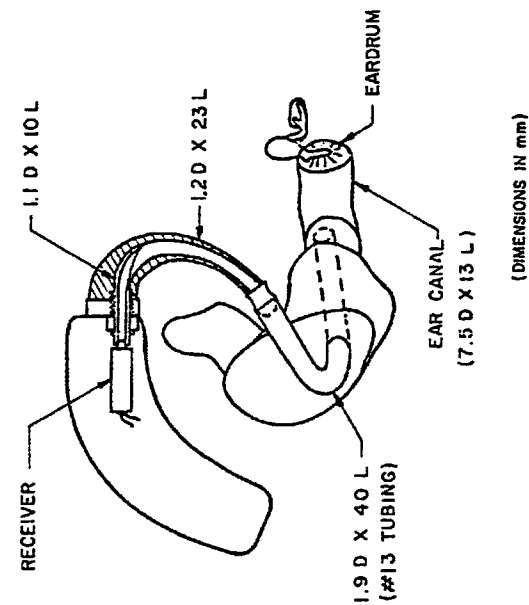
EARPHONE-TO-EARDRUM COUPLING FOR OVER-THE-EAR HEARING AID.

FIGURE 2-4

46
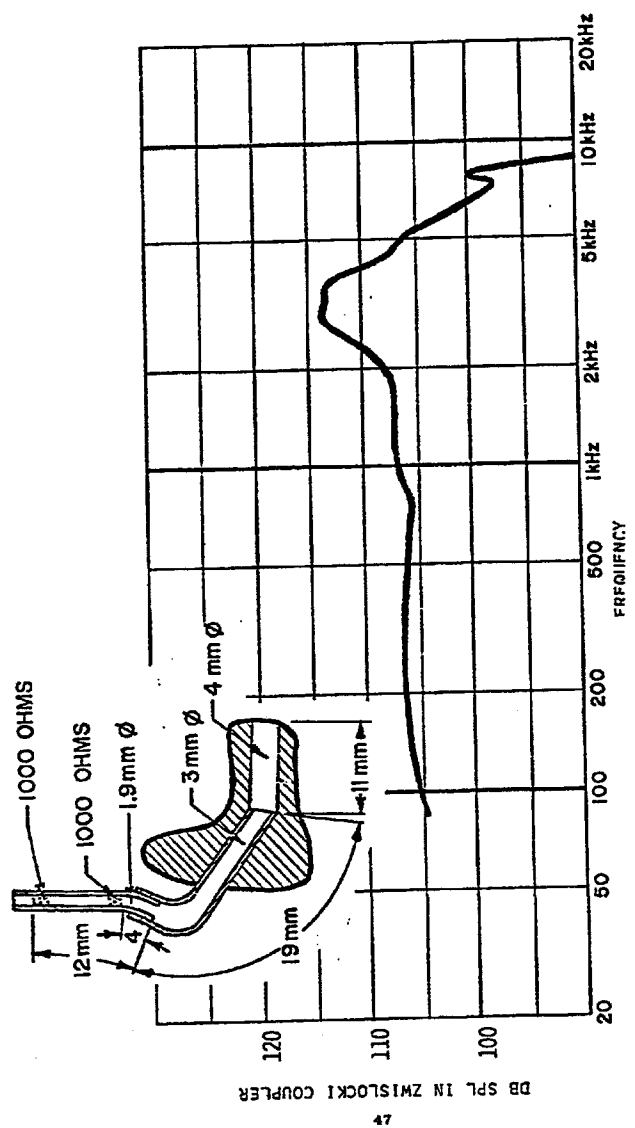
FIGURE 2-5 FREQUENCY RESPONSE OF BP-1712 WIDEBAND EARPHONE WITH "8CR" EARMOLD.

NOTE: EARPHONE MOUNTED IN OTE HEARING AID, AS IN FIGURE 2-4, AND DRIVEN FROM LOW IMPEDANCE AMPLIFIER (.7 mA DC BIAS, .56VAC BEHIND 330 OHMS SOURCE RESISTANCE).

The response of a selected BP-series earphone with this "16KM" (for 16 kHz; Monser) earmold is shown in Figure 2-6.

Unlike the 8CR earmold, the 16KM earmold does not provide a response maximum at 2.7 kHz when used with the BP-series earphone, requiring that the microphone or amplifier response be tailored to compensate for the loss of external ear resonances. (Since the time of this study, a new wideband earphone--Knowles ED-series--has become available. The response of that earphone with the 16KM earmold does exhibit the desired response maximum at 2.7 kHz).

The heavy damping employed in the 16KM earmold may mean impractical battery drains in applications requiring large undistorted outputs (see discussion below). Nonetheless it illustrates that current transducer technology permits 16-kHz bandwidth in hearing aids. At least two individuals appear to be benefiting from this extended bandwidth. Both have an unusual hearing loss configuration; a profound loss at the standard audiometric frequencies but near-normal hearing above 10 kHz (Halperin et al., 1977). A body-worn frequency-up-shifting hearing aid provides electrical drive to a BP-1712 earphone mounted in an earmold shell (with 16KM coupling). In that application, a few extra milliwatts of battery drain was considered a small price to pay for useable output in the 12-18 kHz region. (Preliminary experiments indicate that the new ED-series earphone will allow the construction of ITE aids with acceptable battery drains and 16-kHz bandwidths.)
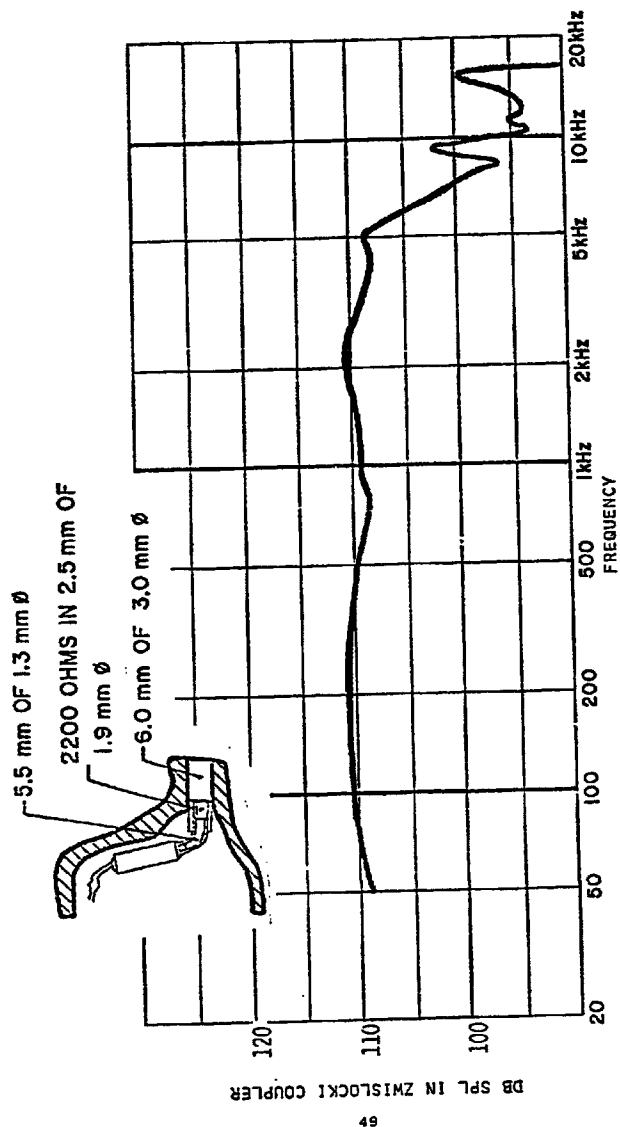
FIGURE 2-6 FREQUENCY RESPONSE OF LOW-IMPEDANCE BP-SERIES EARPHONE WITH "16KM" EARMOLD.

NOTE: EARPHONE MOUNTED AS SHOWN AND DRIVEN FROM LOW IMPEDANCE AMPLIFIER (5 mA DC BIAS, .32VAC BEHIND 70 OHM SOURCE RESISTANCE).

DB SPL IN ZWISLOCKI COUPLER

FREQUENCY

20kHz 10kHz 5kHz 2kHz 1kHz 500 200 100 50 20

120 110 100

5.5 mm OF 1.3 mm Ø
2200 OHMS IN 2.5 mm OF
1.9 mm Ø
6.0 mm OF 3.0 mm Ø

49

C. Response Smoothness

The question of how irregular the frequency response may be before it has a noticeable effect on the fidelity has not been as well studied as the effect of restricting the frequency range.

Bücklein (1962) studied the effect of, among other things, 10 and 20 dB peaks and dips at 3.2 kHz in an otherwise-flat-response transmission system. The peaks and dips that he used appeared fairly sharp, with an apparent 3-dB bandwidth of roughly 10 percent. He found that 100 percent of his observers could detect both the 10 and 20 dB peak, but less than half could detect the 20 dB dip and only 10 percent could recognize the 10 dB dip. Flanagan (1957), in a study of the difference limen for formant amplitude, found that a change of 3 dB in the amplitude of the second formant can be detected approximately 50 percent of the time. From these results, it seems reasonable to infer that a response irregularity in a high fidelity system of approximately 3 dB can be detected under appropriate conditions when the source material is speech.

A procedure based on loudness calculations has been adopted recently by Consumers Union for rating the frequency response accuracy of high fidelity loudspeakers (Anon, 1977). With this procedure, the loudspeakers are driven with a wideband "pink" noise (equal energy in each one-third-octave band) at an overall level adjusted to produce a calculated loudness (Stevens, 1972) of 88 sones, equivalent to a perceived loudness level of 90 PLdB in the 110 to 14,000 Hz frequency range. In each of the 21 one-third-octave bands in that range, the loudness in sones corresponding to the total sound-power output of the

loudspeaker in that band is calculated and compared to the sones

calculated for a theoretically-perfect loudspeaker producing exactly

74 dB SPL in that band. (A 74 dB SPL in each of those 21 bands pro-

duces a calculated loudness of 88 sones.) A percentage accuracy is

then calculated from the deviations from perfection, averaged over the

21 bands. For purposes of the accuracy calculations, the individual-

band deviations from perfection are obtained by subtracting the ratio

of measured to ideal sones from 1.0 (perfection), and then subtracting

the absolute value of that difference from 1.0. Thus a measured power

equivalent to either .5 or 1.5 times the ideal loudness would produce

a 50 percent accuracy score for that band. For example, a loudspeaker

which had a perfectly flat "power response" everywhere except for a

7 dB dropoff (half loudness) in the 125-Hz band would have a calculated

accuracy score of (50 + 20x100)/21 = 98 percent. For our purposes, the

accuracy score described above will be called the "21-band accuracy

score" to distinguish it from a "25-band accuracy score" discussed

below.

The 21-band accuracy scores obtained recently at Consumers Union

on 16 models of "low-priced" ($100 to $200 per pair) high-fidelity

loudspeakers ranged between 63 percent and 93 percent, with a median

value of 80 percent. Listening tests were said to have borne out the

utility of the accuracy scores, although "experience has taught us

that a group of listeners won't readily agree on which of two speakers

is more accurate when the speakers' scores differ by eight points or

less" (Consumers Reports, 1977).

In light of the dearth of published information in this area, a

reasonable goal for the smoothness of the frequency response of a

high-fidelity hearing aid would appear to be a 21-band accuracy score

of 80 percent. Such a hearing aid would have a calculated accuracy

score equal to or better than half the $100-$200/pair loudspeakers

tested recently by Consumers Union. A more stringent goal would be a

21-band accuracy score equal to the median (89%) of the expensive

($600-$1,000 per pair) "State of the Art" loudspeakers tested recently

(Consumers Reports, 1978). The stringency of this latter goal was

evident during the listening-test experiments described below: Simply

relocating a "perfect" pair of loudspeakers in a listening room may

change the one-third-octave response at the listening position suf-

ficiently to reduce a 21-band accuracy rating from 100 percent to 85

to 90 percent.

D. Individual Variation in Ears

If a separate hearing aid were to be designed for each user, it

would presumably be possible to take into account individual eccentri-

cities in external-ear ("earcanal") resonances and eardrum impedances.

To be economically practical, however, a high-fidelity hearing aid

design must be based on average data. Under those circumstances,

individual variations in outer-ear resonance and eardrum impedance may

cause the (insertion) gain and the (insertion-gain) frequency response

of a hearing aid to deviate substantially, for a given individual, from

the design nominals.

An estimate of the individual differences in outer-ear resonance was provided in the data of Filler, Ross and Wiener (1945). In that report, individual sound-field-to-eardrum-pressure curves for 12 male and 2 female subjects were given, which can be compared to the overall average curves for the same subjects shown by Wiener and Ross (1946). The standard deviation of the individual curves (from the average curve) ranged from 1-2 dB below 1800 Hz up to 4-7 dB in the 5 to 8 kHz region with peaks at 2.1 and 3.3 kHz. The peak deviations occurred mostly because individual external-ear resonance frequencies were lower or higher than average. No individual eardrum-pressure-level curve deviated more than 7.5 dB from the average curve below 5 kHz, but the majority deviated by at least 5 dB at some frequency below 5 kHz.

As part of a study leading to a validation of the modified Zwislocki coupler, Sachs and Burkhard (1972) reported the probe-tube-microphone measurement of the sound pressures developed in 11 ears (6 male and 5 female) by subminiature hearing-aid earphones. The standard deviation of the level of pressure developed ranged from approximately 1 dB at 1 kHz to 5 dB in the 6-8 kHz region. Greater pressure levels were developed in female ears (by 3 to 5 dB at the higher frequencies).

While the variations in outer-ear resonance and eardrum impedance are only partially independent variables, it is clear that no hearing aid designed for the average ear can be expected to produce an insertion gain, in the majority of individual cases, which does not have at least one deviation of perhaps 7 dB at the worst frequency. The

subjective importance of such deviations to a long-term hearing aid wearer has not been studied, although it is known that even larger deviations in unaided frequency response can occur due to the accumulation of earwax in the canal; deviations with such a gradual onset that they often go unnoticed by the sufferer until the canal becomes almost completely blocked. In most cases, it thus seems reasonable to assume that satisfactory adaptation to a slightly inaccurate insertion-gain frequency response will make it unnecessary to provide modification for individual eccentricities.

E. Closed Ear Effects

At one time, specifying the airborn stimulus to the auditory system was thought to involve more than simply measuring eardrum pressure because of what Munson and Weiner (1952) called the "missing 6 dB problem" which was attributed by Wever and Lawrence (1954) to a "closed ear effect." More recent experimental studies have indicated that nothing is missing if careful physical- and psycho-acoustic procedures are employed in determining the auditory stimulus. Some of these experiments were summarized by Killion (1978).

There are two closed-ear effects which can occur when the earcanal is sealed off with an earmold, however, which cannot be explained away as experimental artifacts. These are discussed below.

1. Enhanced bone conduction

The increase in the level of bone-conducted sound when the earcanal is occluded is well known, and has been used for nearly 90 years in the otological Bing test for differential diagnosis of conductive

versus "nerve" types of hearing loss. Although a discussion of the mechanism is beyond the scope of this chapter, data obtained by Watson and Gales (1943), Tillman (1962), and informal experiments of the writer indicate roughly 20 dB of low-frequency enhancement can occur when the earcanal is occluded with a well-sealed earmold.

Someone with normal hearing wearing a high-fidelity hearing aid set for unity gain may well find the hearing aid subjectively "disappears" for external sounds. But the enhanced level of autogenously generated sounds (chewing, swallowing, etc.) may make such normally unnoticed sounds quite noticeable. Common background noises--which produce a 20 to 25 dB equivalent hearing loss in a typical residential room (Olson, 1967)--may be loud enough to mask normal-level autogenously generated sounds, but not loud enough to mask enhanced-level sounds.

In some cases, interference with normal communication is possible. The writer has found the crunch of Wheaties so magnified that he occasionally missed parts of breakfast-table conversation while wearing unity-gain ITE aids. (The problem generally disappeared near the end of the bowl as the cereal became sufficiently soggy.)

Someone with the hypothesized cochlear loss discussed in Section I (loss of sensitivity for low-level sounds but sensibly normal hearing for high-level sounds) may or may not hear autogenously-generated noises. If he does, the gain required of the aid for low-level airborne sounds will presumably be sufficient to restore the normal balance between airborne and low-level autogenously-generated sounds.

To the writer's knowledge, no data are available regarding the importance of autogenous noises to the hearing-aid wearer. It is

interesting to speculate that the oft-reported preference for vented or "open" earmolds (especially among users with normal low-frequency hearing) may be related to the enhanced bone conduction which results when a well-sealed earmold is used.

2. Reduced acoustic-reflex attenuation

The most recent estimate of the effect of the acoustic reflex on sound transmission in the human is that some 10 dB of attenuation is provided for low-frequency sounds, but only minimal attenuation is provided for high-frequency sounds (Rabinovitz, 1976). Thus the subjective result of the activation of the acoustic reflex is presumably a reduced "bass response". While the acoustic reflex will have negligible effect on the eardrum pressure produced by a free-field source, it will increase the eardrum pressure produced by a hearing-aid earphone coupled to an occluding earmold. In the latter case, the acoustic source impedance presented by the hearing-aid earphone is so high that the eardrum pressure will be nearly proportional to the total load impedance of the earcanal volume and eardrum impedance. (Indeed, this is the basis of the clinical measurement of the acoustic reflex). Thus while exactly the same attenuation in low-frequency transmission can be expected whether the earcanal is open or occluded, Rabinovitz's data indicate the total impedance change may be on the order of 20 to 40 percent, corresponding to a 1 to 3 dB change in eardrum pressure at low frequencies. This is a small enough change that the difference in subjective bass response for very-high-level (reflex activating) sounds may be of academic interest only.

III. AMPLITUDE RESPONSE REQUIREMENTS

In the amplitude domain, one generally thinks of the dynamic range between maximum undistorted output levels and noise levels, non-linear (harmonic and intermodulation) distortion, and Automatic Gain Control (AGC) functions. Some performance requirements imposed on a hearing aid if it is to provide high fidelity sound reproduction are discussed below. As before, examples are provided to illustrate practical hearing aid designs meeting these requirements.

A. Peak Output Levels

There is no easy answer to the question of what the maximum un-distorted output of a wearable unity-gain-sound-reproduction system should be. If the peak outputs of rock music played at a discotheque are to be reproduced, a 130 dB SPL capability may be required (R.W. Peters, personal communication). A similar capability would be re-quired to reproduce the peak levels produced by some aircraft and industrial noise sources. As summarized by Miller (1974), however, such levels are hazardous to the hearing mechanism.

1. Speech and everyday sounds

In everyday conversational settings, the highest levels at the hearing aid microphone will normally be generated by the user's own speech. The data of Dunn and Farnsworth (1939) indicate that the overall speech levels measured at the talker's ear were about equal to the levels 30 cm in front of his lips. Thus, the Dunn and White (1940) data on instantaneous peak levels in speech measured at 30 cm

may be used directly. For normal conversational speech, instantaneous speech peaks of 90 to 95 dB SPL occur with some regularity (in 1% to 5% of 1/8th-second intervals). (Due to the head-shadow effect, the high frequencies will be attenuated somewhat, but since the majority of the peak energy lies below 1000 Hz, this will have little effect on the overall peak levels.) Unless the conversation becomes agitated, there-fore, the instantaneous peak levels at the microphone will be 90 to 95 dB SPL; a range which has often been used as a minimum-undistorted-input-level goal in hearing aid design. If the conversation becomes agitated, a shout or forceful expletive can easily produce 100 to 105 dB instantaneous peak SPL at the microphone of the speaker's own hearing aid; as can a child's enthusiastic greeting. (These numbers are actually quite conservative: The writer's young daughter can pro-duce a 114 dB instantaneous peak SPL "Hi, Dad" at $2\frac{1}{2}$ feet.)

Other commonly encountered sounds such as the clack of a type-writer key or a finger snap at arm's length can also produce a 100 to 110 dB instantaneous peak SPL. A spoon dropped onto a plate can produce a 110 to 115 dB instantaneous peak SPL.

2. Live music

In live performances of classical music, Marsh (1975) reported that in a good main-floor seat in Chicago's Orchestra Hall ..."a fully scored orchestral passage in a Mendelssohn Symphony reaches approxi-mately 95 dB on a decibel meter". Marsh reported that approximately the same levels are reached during a similar passage in the front benches at Grant Park or at the edge of the Ravinia stage. The

typical instantaneous peak factor for an orchestral passage of this sort is 5 to 10 dB, indicating instantaneous peaks of 100 to 105 dB SPL at these three Chicago-area locations. Similarly, for a typical listening position in a large music hall, Olson (1967) reported the instantaneous peak sound pressure level as 100 dB SPL.

These more recent data are consistent with earlier measurements made of the Philadelphia Orchestra playing a wide variety of selections during a three hour recording session (Fletcher, 1942). At the measurement location, 20 feet from the center of the orchestra (!), the instantaneous peak levels were estimated at 112 dB SPL.

All things considered, an undistorted input capability of 105 dB instantaneous peak SPL--referred to the sound field--appears to be a reasonable minimum requirement for a high-fidelity hearing aid.

What is needed for our present purposes, however, is information on the peak output levels required of the hearing aid earphone. This requires consideration of the frequency distribution of peak levels in music as well as consideration of the increased eardrum pressure levels normally produced by external-ear resonances.

The frequency distribution of the maximum instantaneous peak levels for a 75-piece orchestra was given by Fletcher (1931) and is shown in Figure 2-7. Several more recent measurements have indicated that the high-frequency peak levels do not fall off as fast as Fletcher's curve indicated (and indeed Fletcher himself intentionally excluded the cymbols; an instrument whose inclusion would have brought the 10 kHz level back up near the level at 250 Hz). It is sometimes argued, therefore, that a no-compromise system would allow no drop
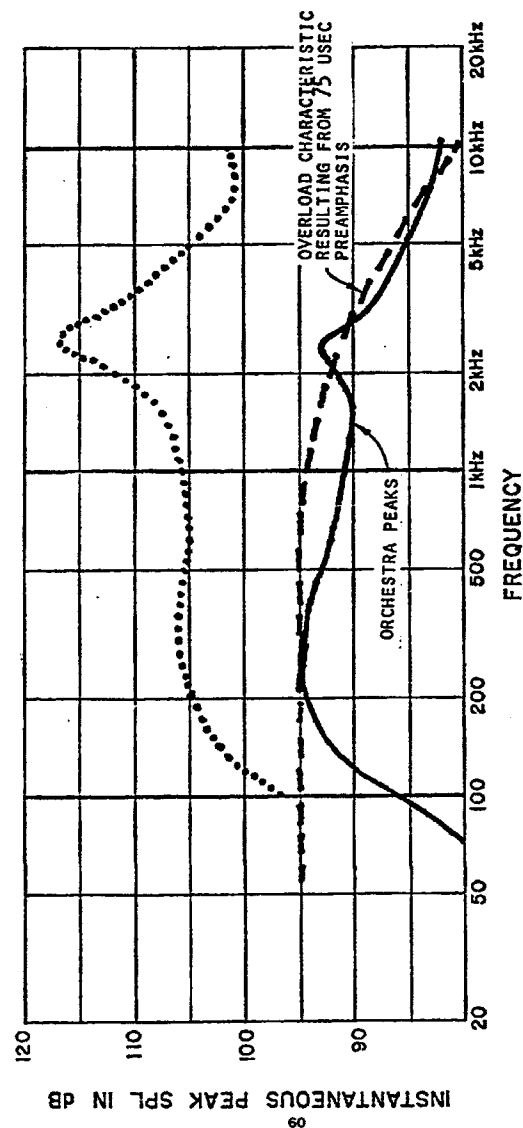
FIGURE 2-7 MAXIMUM PEAKS PRODUCED IN HALF-OCTAVE BANDS BY LARGE ORCHESTRA, REFERRED TO AUDIENCE AREA (——), INSTANTANEOUS PEAK (OF SINE WAVE) EARDRUM PRESSURE CAPABILITIES REQUIRED OF HEARING AID TO REPRODUCE LARGE ORCHESTRA (••••),

NOTE: ORCHESTRA PEAKS CORRESPOND TO INSTANTANEOUS PEAKS OF 105 dB SPL OR PEAK READINGS OF 95 dB SPL ON SOUND LEVEL METER.

in the high-frequency peak capability.

Since occasional high-frequency overload is generally found to be less annoying than a constant hiss, however, a "75 microsecond" high-frequency preemphasis (a 6 dB/octave rollup above a 2.1 kHz corner frequency) remains the standard for FM broadcasting. Similarly, current AES standards on prerecorded tape and phonograph records call for a preemphasis ranging from 150 to 75 microseconds. The effect of the preemphasis is to improve the signal-to-noise ratio of recorded material at the expense of a reduced high-frequency overload capability (which falls off at 6 dB per octave above 2100 Hz).

Not surprisingly, the frequency dependence of the instantaneous peak-pressure requirements is not much different whether based on the Fletcher curve or the assumption of a 75 microsecond preemphasis. This comparison is shown in Figure 2-7.

In terms of system requirements, approximately 10 dB of "headroom" is required in any half-octave band over the instantaneous peak levels shown in Figure 2-7. This comes about because the presence of energy in other frequency bands can, when added to that present in a given band, produce instantaneous peaks greater (by approximately 10 dB in the case of half-octave bands) than the peak which would be produced by the in-band energy acting alone. Looked at in terms of waveforms, the in-band peak can ride on top of a wideband waveform. The measured instantaneous peak overload capability of a system at a given frequency (measured as the instantaneous peak of a just-unclipped sinewave signal) must therefore exceed the half-octave-band instantaneous peak measurements of music by approximately 10 dB.

The output requirements for a hearing aid must also take into account the increase in eardrum pressure produced by outer-ear resonances. On a random-incidence basis (as in a concert hall where sounds are arriving from all directions), this increase amounts to approximately 15 dB at 2700 Hz (Shaw, 1976). When combined with the considerations of the previous paragraph, this means a hearing aid must be capable of producing an instantaneous peak eardrum-pressure level at 2700 Hz which is 25 dB greater than the peak level shown in Figure 2-7. The upper (dotted) curve in Figure 2-7 shows the estimated instantaneous peak (of a sinewave) output requirements for a high-fidelity hearing aid operating at unity insertion gain for high-level signals.[1]

3. Peak output capability of hearing-aid earphones

As discussed elsewhere (Killion, 1978a), the maximum undistorted output available from a subminiature hearing aid earphone is generally determined by the overload of the hearing aid amplifier rather than overload of the earphone itself. This is certainly true of Class A amplifiers biased at the .5 to 2 mA of battery drain which has found market acceptance in OTE hearing aid design.

Figure 2-8 shows the maximum undistorted output of a BP-1712 wideband earphone when mounted in an OTE hearing aid as shown in Figure 2-4

---

[1] Entirely different considerations apply when output limitation is used in high-gain hearing aids to prevent user discomfort. In that application, the maximum output of the hearing aid should probably not have a peak at 2.7 kHz.
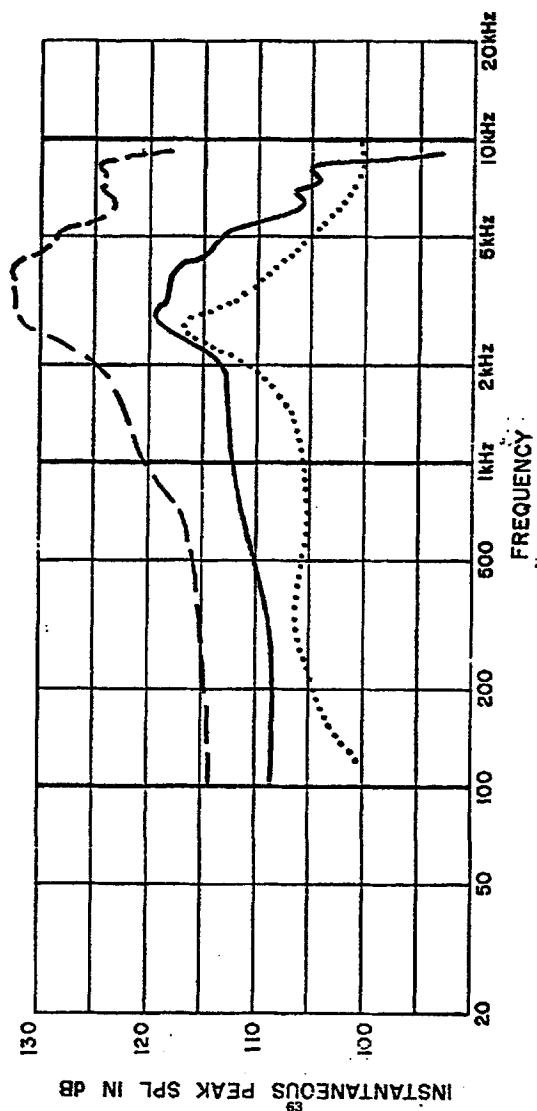
FIGURE 2-8   PEAK EARDRUM PRESSURE REQUIREMENTS (•••••) COMPARED TO MAXIMUM LINEAR
OUTPUT OF BP-1712 EARPHONE WITH 8CR EARMOLD, AS LIMITED BY:

EARPHONE OVERLOAD (———)

CLIPPING IN .7 mA EXPERIMENTAL AMPLIFIER (———)

63

and coupled to the 8CR earmold. The maximum undistorted output determined by earphone overload (dashed curve) substantially exceeds the limitation imposed by typical amplifier overload (solid curve). The experimental amplifier chosen for this comparison was biased at .7 mA of battery drain; a battery drain sufficient to meet the instantaneous peak output requirements shown in Figure 2-7 and reproduced as the dotted curve in this Figure. Higher output levels would have been obtained with greater battery drain and lower earphone impedance, and vice versa.

Due to the heavy damping used in the 16KM earmold, a lower earphone impedance and a greater amplifier battery drain is required to drive the BP-series earphone to the instantaneous peak output requirements shown in Figure 2-7 when the earphone is coupled to a 16KM earmold. Figure 2-9 shows the maximum undistorted output of a 150 ohm BP-series earphone driven by a Class A amplifier biased at 5 mA battery drain. The latter is roughly ten times the battery drain levels which have found market acceptance in practical In-The-Ear hearing aids. (Preliminary experiments using the 16KM earmold with the new Knowles ED-series earphone indicate the instantaneous peak output requirements shown in Figure 2-7 may now be achievable with battery drains practical for ITE hearing aids.)

B. Input Noise Level

The input noise level of a high fidelity hearing aid should be less than that of the ambient noise levels likely to be encountered by a user. The A-weighted noise level during a quiet listening period in
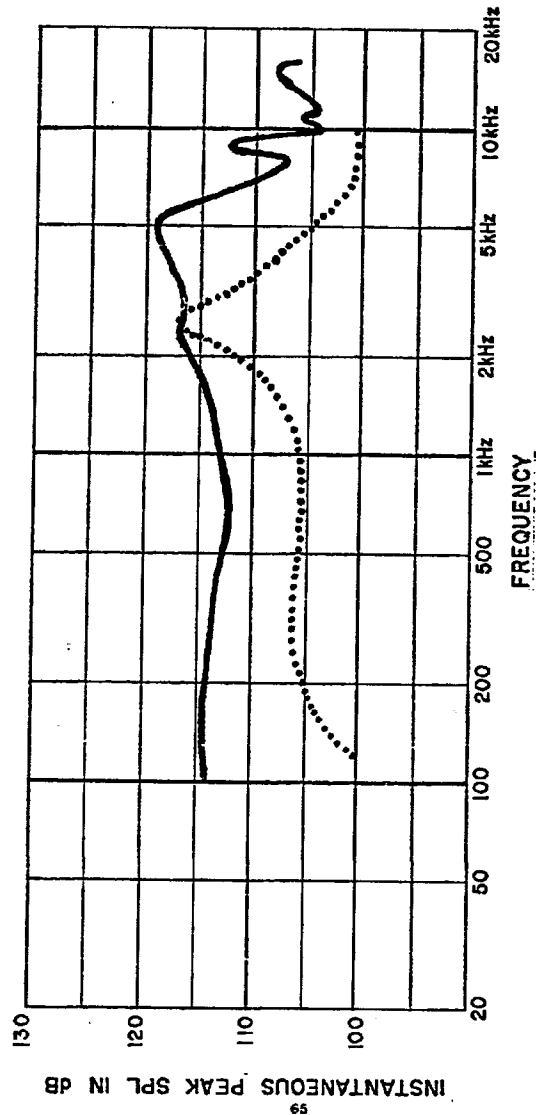
FIGURE 2-9 PEAK EARDRUM PRESSURE REQUIREMENTS (••••) COMPARED TO MAXIMUM LINEAR OUTPUT OF LOW-IMPEDANCE BP-SERIES EARPHONE WITH 16 KM EARMOLD AND 5 mA EXPERIMENTAL AMPLIFIER (———).

a theater or auditorium may drop to 32 dB (Fletcher, 1942; Olson, 1967). Residential noise levels are generally higher. Seacord (1940) measured noise levels in a large number of residential rooms and found an average level of 43 dB A-weighted; 90 percent of the levels fell between 33 and 52 dB(A). Seacord's data have found common acceptance, although it is generally believed that the greater use of forced-air heating and air-con:::tioning systems has acted to increase average levels since his data were obtained.

The input noise level in a modern hearing aid is determined almost entirely by the microphone noise level. Currently available subminiature microphones have typical A-weighted noise levels equivalent to a 26 dB SPL ambient noise level. Even in a quiet auditorium, therefore, the microphone noise level would add less than 1 dB to the apparent A-weighted ambient level. The aided threshold determined by typical microphone noise levels is within a few dB of normal threshold and can be better than normal with special design (Killion, 1976a).

C. Distortion

We will use the term distortion to mean "nonlinear" distortion in the restricted sense of a distortion which results in the generation of new frequencies appearing in the output but not present in the input stimulus. An example of nonlinear distortion would be the overload of an amplifier so that clipping occurs. Common measures of nonlinear distortion are total harmonic distortion and intermodulation distortion, both of which measure the relative strength of the new frequencies created by the nonlinearity in the presence of one or more sine-wave input frequencies.

In a study of the amounts of distortion tolerable in a high-fidelity system, Olson (1957) found that total harmonic distortion levels of approximately one percent were just detectable—and total harmonic distortion levels below three percent were not considered objectionable—in a system with an 8-kHz upper cutoff frequency. Two systems were tested, one using a single-ended triode amplifier and one using a single-ended pentode amplifier. In both cases, the distortion increased with increasing output and the distortion percentages correspond to the total harmonic distortion measured with a (sine-wave) output level equal to the peak levels encountered in the music used as program material.

Nonlinearities in the system which occur near the upper cutoff frequency are difficult to examine using harmonic distortion measures, because the harmonic distortion products (at $2f_1$, $3f_1$, etc.) may lie above the cutoff frequency. High-frequency distortion can be easily detected, however, by applying two high-frequency sine waves to a system and measuring the system output at the <u>difference</u> frequency (the "CCIF method" for measuring intermodulation distortion).

Tests made by the British Post Office and reported by Moir (1958) indicated that CCIF intermodulation "distortion on speech and music is not detectable when the quadratic or cubic difference tones are... below one percent..." Termin and Pettit (1952) reported that CCIF intermodulation distortion becomes objectionable at a value of three percent to four percent when the difference frequency lies in the 400 to 5000 Hz range.

More recent tests employing "golden ears" have produced similar results, as summarized by Milner (1977) and Davis (1978). Distortion levels below two percent are generally inaudible on musical material, and even gross distortion levels (6 to 12%) are sometimes inaudible. In general, the just-audible distortion levels for musical material are at least ten times greater than the just-audible pure-tone distortion levels.

One complication which arises when attempting to apply these results to hearing aids is the necessity to translate the results into eardrum or coupler-pressure levels. Although eardrum pressure and field pressure are roughly equal below 1000 Hz, the combined effect of head diffraction and external-ear resonance results in a 10 to 20 dB boost in eardrum pressure levels in the 2000 to 5000 Hz region (Wiener and Ross, 1946). One consequence of the treble boost provided by the head and outer ear is that the ear is able to detect lower levels of harmonic distortion (when the harmonics fall in the 2000 to 5000 Hz region) than would otherwise be audible. (In addition, small head movements may place one ear in a "null" position for the fundamental tone in the listening room, which can result in incredibly low pure-tone distortion detection levels.)

In recent experiments where the listeners employed headphones, Gabrielsson <u>et al.</u> (1976) studied the audibility of quadratic and cubic "power series" types of distortion. They concluded that 2.5 percent quadratic and 1.25 percent cubic distortion levels would be rarely detected, even on relatively simple flute or clarinet tone stimuli. The allowable levels on music and speech were two to three times higher.

A further complication is the level-dependence of the ear's sensitivity to distortion. At low sound pressure levels, the level of any harmonic or intermodulation distortion products may lie below the normal threshold of hearing. At high levels, the increased upward spread of masking (Wegel and Lane, 1924) and the distortion of the ear itself may mask externally generated distortion products. Thus no single-number distortion specification will apply at all listening levels.

The tests reported by Olson (1957) were carried out in a small listening room at a level of "about 70 dB". Assuming a similarity in this level to the 75 dB peak levels measured on a sound level meter Olson reported for other similar listening tests, this corresponds to instantaneous peak levels of approximately 85 dB SPL; only slightly greater than the 70 to 80 dB SPL which a study of the masking literature indicates is the region where the ear is most sensitive to distortion. Thus, the values for detectable and tolerable distortion obtained by Olson and others can reasonably be applied as a requirement for a high-fidelity hearing aid only for output levels of perhaps 50 to 90 dB SPL (measured at the eardrum or in an ear simulator). At lower and higher levels, some relaxation is clearly in order, to perhaps 10 percent at 30 dB and 110 dB eardrum SPL.

The last problem in attempting to arrive at reasonable distortion specifications is probably the most important one. Unless the precise distortion mechanism (peak clipping, center clipping, curved transfer characteristic, etc.) is understood, no single distortion measurement can provide reliable information as to how clean the sound will be to

a human listener. Thus, one percent of "soft peak clipping" (as employed in the determinations discussed above) may be inaudible, but one percent of center-clipping distortion may be intolerable!

Similar difficulties are encountered in attempting to study the effect of distortion on speech discrimination. Thus, Peters and Burkhard (1968) found that the 40 percent total harmonic distortion produced by one system had negligible effect on speech discrimination, whereas another system whose distortion measured only 20 percent resulted in a loss of 40 percentage units in speech discrimination score.

In light of the available information, a reasonable initial goal for a high-fidelity hearing aid with an 8000 Hz bandwidth would appear to be a maximum total harmonic distortion or CCIF intermodulation distortion of two percent for output levels between 50 and 90 dB eardrum SPL, assuming the distortion mechanism is a simple one, with a linear relaxation to 10 percent at 30 and 110 dB SPL. Figure 2-10 illustrates such a requirement. Listening tests should be employed as a final check.

1. The distortion levels in ears

A somewhat different approach to the problem of defining allowable distortion levels as a function of overall SPL is to estimate the apparent distortion produced in the normal ear itself as a function of overall eardrum SPL's. (We will assume that the distortion in the impaired ear is at least as great.) One such estimate is shown in Figure 2-11 which represents an attempt to pull together some two dozen studies on aural distortion and combination tones after converting the available data to estimated eardrum-pressure levels based on the
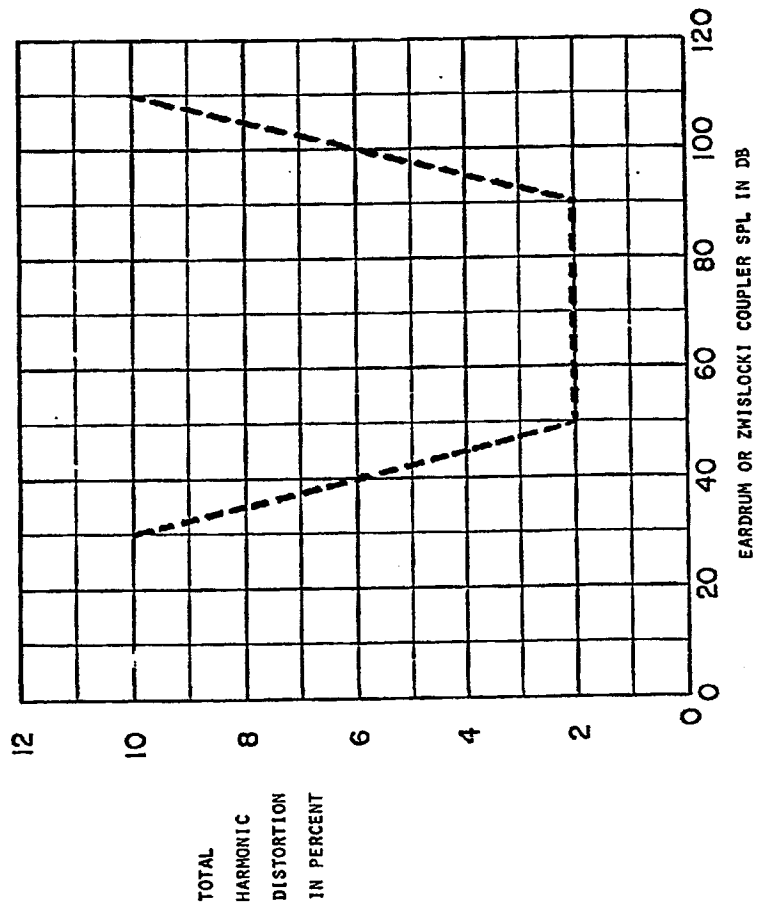
FIGURE 2-10   ESTIMATED MAXIMUM HEARING-AID DISTORTION WHICH WILL BE INAUDIBLE ON SPEECH OR MUSIC

71
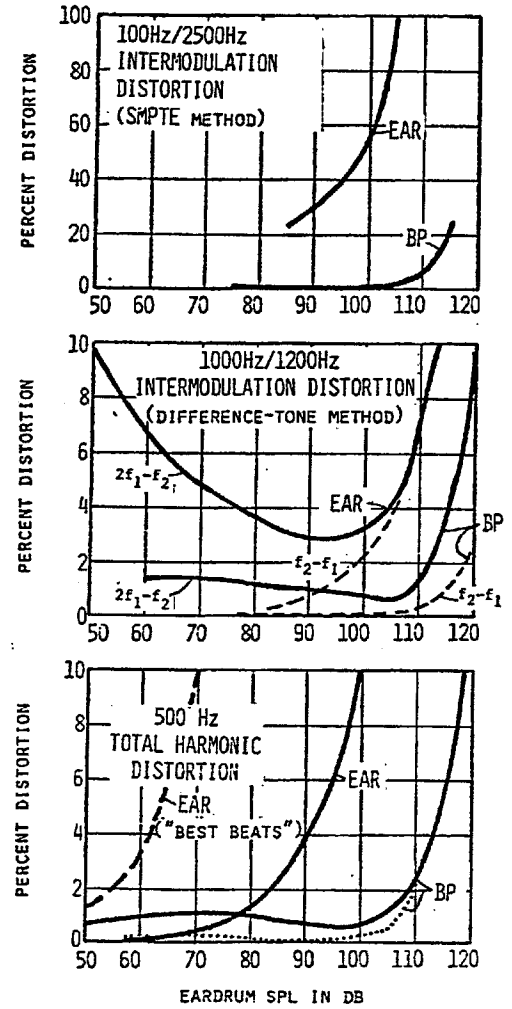
FIGURE 2-11   APPARENT DISTORTION LEVELS IN NORMAL EARS AND IN
THE BP-SERIES EARPHONES.

72

conversions summarized by Killion (1978).

The upper graph in Figure 2-11 shows an aural "SMPTE intermodulate distortion" (low-tone/high-tone method) estimate based on recent data obtained by Zwicker (1976). Zwicker was looking for the psychoacoustic equivalent of the period histogram typically found in single-unit neural firings recorded from the eighth nerve. Zwicker found that a brief 2500 Hz tone pip was made completely inaudible when added to an intense 100 Hz tone on one-half of the 100 Hz wave-form, but quite audible on the other half.

The middle graph shows an aural "CCIF intermodulation distortion" (difference tones) estimate based on the average of the results of several studies on the level of the two most prominent combination tones: the cubic difference tone at $2f_1 - f_2$ (Zwicker, 1955; Goldstein, 1967; Helle, 1969; Hall, 1972) and the simple difference tone at the frequency $f_2 - f_1$ (Bekesy, 1934; Speith, 1957; Zwicker, 1955, 1958; Wenner, 1968; Hall, 1972). These combination tones are easily heard, especially while playing a musical recorder duet at close range!

A good estimate of the harmonic distortion levels in the ear is the hardest to come by. For years after it was first used by Wegel and Lane (1924), the "Best Beats" method was the commonly accepted method for estimating the level of aural harmonics (Bekesy, 1934; Moe, 1942; Egan and Klumpf, 1951). The dashed curve in the bottom graph is derived from several studies in which that method was used. The problem with that method, as pointed out 40 years ago by Trimmer and Firestone (1937), is that no one ever _hears_ these harmonic distortion

products. A more conservative estimate is shown by the solid curve labeled EAR in the lower graph, based on a "phase" method used by Clack et al. (1972) and deBoer and Bouwmeester (1975).

2. The distortion levels in a wideband earphone

The results of distortion measurements on a BP-1712 wideband earphone are also shown in Figure 2-11. These measurements were obtained with the earphone mounted in an OTE hearing aid case, as shown in Figure 2-4, and coupled to a Zwislocki coupler through a well-damped "6R12" earmold (Knowles and Killion, 1978). The earphone was driven from a high-impedance source which, in combination with the high-frequency emphasis provided by the 6R12 earmold, was expected to provide a worst-case distortion estimate. Use of a low-impedance source reduces the earphone distortion, as shown in the dotted curve in the bottom graph of Figure 2-11.

The conclusion which can be drawn from the comparison of Figure 2-11 is that the distortion produced by available wideband hearing aid earphones is typically well below that generated by the nonlinearities in the ear. Some 20 dB of pure conductive hearing loss would be required before the earphone distortion became comparable to that of the ear.

An informal listening test (with a 6R12 earmold) tended to confirm these conclusions. With pure-tone stimuli at optimum levels, harmonic distortion was clearly audible with high-impedance electrical drive to the earphone, but generally not audible under low-impedance drive conditions. With music or speech, earphone distortion was generally inaudible with either drive condition.

In a practical hearing aid design, therefore, the important limitation will generally be amplifier distortion. Fortunately, amplifier distortion can easily be reduced well below audible levels by the use of negative feedback. A comparison between the lower dotted curve in Figure 2-11 and the estimated maximum-inaudible-distortion curve of Figure 2-10 reveals that a BP-series earphone driven by a low-distortion, low-impedance amplifier will have a total harmonic distortion which is less than one fourth the estimated audible limits.

D. Automatic Gain Control Characteristics

Recall that a unity-gain sound-reproduction system is desired for high-level sounds, coupled with sufficient gain for (desirable) low-level sounds to make them audible to the hearing aid user.

The low-level gain required of a high-fidelity hearing aid can be readily estimated. Although a gain numerically equal to the user's hearing loss would be required to restore his threshold down to audiometric zero levels, such a large amount of gain is commonly found to be unacceptable (Martin, 1973). This is easily understood. Under most circumstances, the masking produced by the background noise levels commonly encountered in residences, offices, etc., render even those with unusually acute hearing incapable of detecting sounds which are less than 15 to 30 dB above commonly accepted audiometric zero levels.

Based on the average room noise and spectra data of Seacord (1940) and Hoth (1941), Killion and Studebaker (1978) estimated that the masking effect of typical residential room noise produces a nearly uniform 23-dB hearing loss across the 250 to 4000 Hz speech frequencies.

(Similar estimates had been made by others. Thus Olson (1957) calculated a 20 to 22 dB loss using slightly different assumptions.) Gain above that required to make background noises audible will be "empty gain" which makes everything louder but does not improve the detection of quiet sounds. Similarly, the upper limit of what is commonly accepted as the range of normal hearing is set at 25 dB above audiometric zero levels. Thus a maximum gain sufficient to reduce the effective hearing loss to 15 or 20 dB has been generally found appropriate.

The input level at which the gain should be reduced to unity can also be estimated. Examination of clinical data and the literature on recruitment indicates that recruitment is typically complete (loudness sens-. is essentially normal) for sounds corresponding to a hearing lev(80 dB or greater. More recently, Barfod (1978) has shown that a nearly linear relationship can exist between the degree of hearing loss and the Hearing Level (HL) at which recruitment is complete. For Barfod's subjects--all of whom had steeply-sloping high-frequency losses with nearly-normal low-frequency hearing--all hearing losses below 50 dB HL were accompanied by complete recruitment above 75 dB HL.

As a practical example, assume a user has a 45 dB HL cochlear impairment with complete recruitment for sounds above 80 dB HL. By our assumptions, he requires a maximum gain of 30 dB for sounds at 15 to 20 dB HL, and unity gain for sounds at 80 dB HL and above. (For speech sounds, 80 dB Hearing Level corresponds to a sound pressure level (SPL) in a 0° in .dence sound field of approximately 95 dB. For pure tones, 80 dB Hearing Level corresponds to approximately 80 to 90 dB SPL in the sound field in the frequency range important for speech perceptions.)

1. Speculations regarding optimum AGC characteristics

In order to avoid constant adjustments of the volume control, an Automatic Gain Control (AGC) system is required. In order to introduce the minimum degradation in perceived sound quality, the operation of the AGC system must be unobtrusive. Compression amplification has found wide acceptance in the broadcast and recording industry for such purposes (Blesser and Ives, 1972). The idea of applying compression amplification to hearing aids is an old one, of course, dating back at least to Steinberg and Gardner (1937). /As originally defined at Bell Telephone Laboratories (Mathes and Wright, 1934), compression amplification meant what is now sometimes called logarithmic compression (to distinguish it from some of the misuses of the term); i.e., a constant ratio between the logarithms of the input and output signal amplitudes. When input and output levels are expressed in dB, for example, a compression ratio of 2:1 corresponds to a 5 dB increase in output level for each 10 dB increase in input level./ Wide-dynamic-range compression amplification was apparently first reduced to commercial practice in wearable hearing aid design by Goldberg (1960, 1966).

Figure 2-12 illustrates one possible input-output characteristic for a hearing aid intended to meet the requirements of the example discussed above. There are four stages of amplification illustrated in Figure 2-12: a low-level constant-gain stage, a mid-level constant-compression-ratio (2:1) stage, a high-level unity-gain stage, and a very-high-level compression limiting stage.

The final compression limiting stage requires further comment. Fast-acting low-distortion compression limiting applied to the
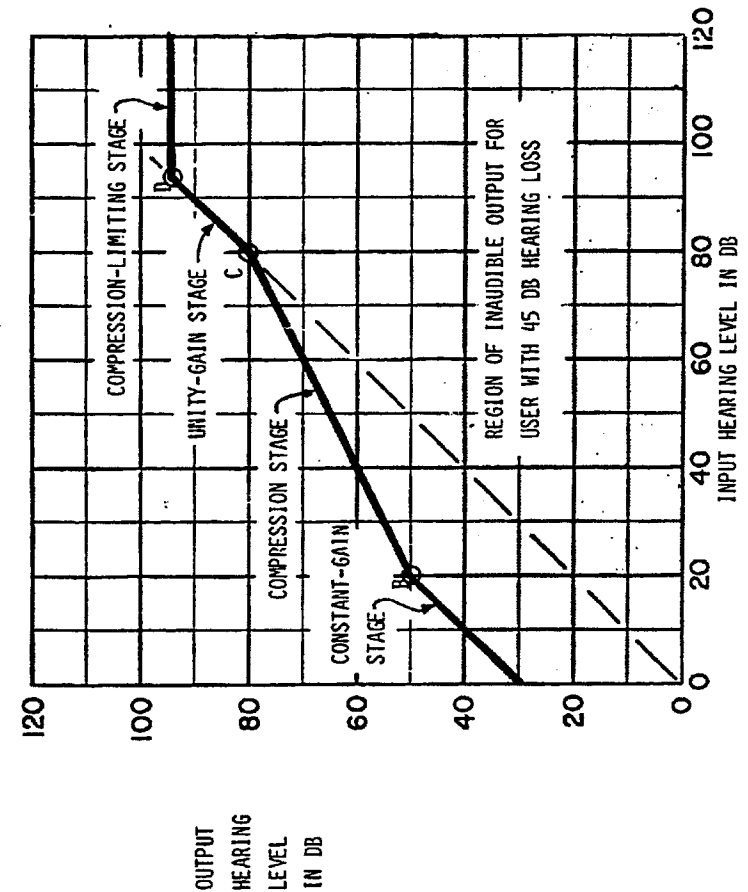
FIG. 2-12 PRESUMED IDEAL APPROACH TO DYNAMIC-RANGE REDUCTION: FOUR-STAGE COMPRESSION AMPLIFIER

78

microphone output for sounds above roughly 100 dB hearing levels (110 to 115 dB SPL) serves to prevent audible distortion or bias shifts in the amplifier when its output capability would otherwise be exceeded. Output limiting should not be required to prevent sounds from becoming uncomfortably loud if the hearing aid has unity acoustical gain for high-level sounds; the user would be exposed to uncomfortably-loud sounds no more often with his hearing aid than without it. Even with unity gain in a hearing aid, however, the writer has observed that the wideband "spectral splatter" accompanying amplifier overload can make an otherwise innocuous sound uncomfortably loud. (One wag has suggested that the obvious solution to hearing aid distortion was for the user to carry a 200 watt amplifier around with him, which would avoid the latter problem.)

Just as the idea of applying compression amplification to hearing aids is an old one, the characteristics illustrated in Figure 2-12 are similar in one regard or another to those suggested by Goldberg (1960, 1966, 1972), Martin (1971), and Villchur (1973). Barfod (1978) has recently presented an argument for a slightly different set of characteristics (which take better account of the relationship he has found between the degree of hearing loss and the hearing level at which recruitment is complete).

Many commercial hearing aids currently employ low-distortion compression amplification with input-output characteristics similar to at least the first two stages shown in Figure 2-12, but to the writer's knowledge, there is little evidence to support one characteristic over another from the standpoint of sound quality as perceived by those with

hearing impairment.

The determination of the optimum AGC input-output characteristics was considered beyond the scope of the present study.

IV. TIME RESPONSE REQUIREMENTS

In this section, the transient response of a hearing aid is considered.

A. Transient Waveform Response

There are two common interpretations for the term "poor transient response". One is the difference between the output and input waveforms viewed on an oscilloscope screen (transient waveform response) when a transient is applied to a sound reproduction system. In general, this difference is an inevitable consequence of any frequency response shaping in the system under test. By the above definition, for example, the ear itself has a very poor transient response because of the resonances in the external ear.

To a reasonable first approximation, a hearing aid system can be represented as a minimum-phase network. Under those circumstances, the transient response can be predicted directly from the frequency response. The frequency response tailoring of a particular hearing aid may or may not be useful, but its effect on "transient response" is inevitable. Indeed, the inverse procedure--obtaining the frequency response of a hearing aid by analyzing its transient waveform response to short clicks--is sometimes used (Studebaker, 1977).

Two papers should be read by anyone interested in "poor transient response". A comprehensive set of frequency-response curves and their corresponding transient-response waveforms was given in a paper by Mott (1944). The surprisingly large changes in waveform due to phase shifts which are nonetheless completely inaudible have often been discussed. Bauer (1974) describes some of the more recent experiments along those lines, experiments indicating that even 90-degree-per-octave rates of phase shifts are generally inaudible.

B. Transient Amplifier Overload

The other type of "poor transient response" generally involves an amplifier which exhibits a slow recovery from overload. As mentioned above, instantaneous peak sound pressure levels of 110 to 115 dB at the hearing aid input are not uncommon. Such peaks can easily cause sufficient amplifier overload to upset the bias levels on the internal coupling capacitors, causing a high-amplifier-distortion condition lasting much longer than the transient itself (Ingelstram, Johanson, Pettersson, and Sjögren, 1971; Killion, Carlson, and Burkhard, 1970). This "blocking distortion" was more often a problem with older amplifier designs, and is much less of a problem now that the majority of hearing aids use some form of fast-acting compression limiting or AGC system.

C. AGC Time Constants

The attack and release time constants of the AGC system used to obtain the desired input-output characteristics throughout the operating range are an important consideration in any sound processing system.

As extensively discussed by Lippman (1978), the proper choice of time constants depends a great deal on the goal set for the AGC system of the hearing aid. When the goal is to maximize speech discrimination, for example, the recent results of Ahren et al. (1977) and Schweitzer and Causey (1977) indicate the attack time should be as short as possible and the release time should be less than 100 msec and perhaps greater than 30 msec.

When the goal is to maximize sound quality, on the other hand, the situation is much less clear. Even under ideal conditions such as found in professional recording studios, the optimum choice of attack and release time for minimum perceived distortion is highly dependent on the program material. Thus, any choice will be "wrong" at least part of the time. These issues were discussed at some length by Blesser and Ives (1972), who reported that values of 10 msec and 150 msec for attack and release times, respectively, have found common acceptance in equipment designed for the broadcast industry. In the absence of reliable research findings on the optimum values for hearing aids (with sound quality as the goal), these values would presumably represent a reasonable first choice for the AGC system of a high-fidelity hearing aid.

When the goal is to prevent audible distortion due to amplifier overload, on the other hand, the attack time must be as short as possible. Generally speaking, a separate compression-limiting system will be required for amplifier overload prevention. If the compression limiting is expected to operate only for occasional

transient peaks, the release time can also be made quite short--perhaps 50 msec or so--since the resulting distortion of low-frequency signals will be so brief as to be unnoticed.

V. EXPERIMENTAL HEARING AID DESIGN EXAMPLES

The objective performance requirements for high fidelity were outlined in the preceding section. Since each individual requirement could be met using existing transducers, the next logical step was to "put it all together" in experimental high-fidelity hearing aid designs. Two such experimental designs were assembled (as binaural pairs): a pair of Over-The-Ear (OTE) hearing aids with 8 kHz bandwidth, and a pair of In-The-Ear (ITE) hearing aids with 16 kHz bandwidth.

The mounting and frequency responses of the microphones and earphone-earmold combinations used in the experimental hearing aids are described in the first two sections below. Construction details for both the OTE and ITE hearing aids are given in each section. An additional pair of In-The-Concha-microphone-pickup hearing aids was subsequently assembled; these are described briefly in a separate section.

The amplifiers used with the experimental hearing aids were of "breadboard" construction mounted in pocket-size cases. Reducing a discrete-component breadboard amplifier to subminiature dimensions is a feat regularly accomplished by hearing aid designers, and was not considered an important part of the present investigation. The electrical characteristics of the experimental amplifiers were an important feature, however, and are described in a third section.

In two final sections, data on the frequency response and nonlinear distortion of the completed aids is discussed. In particular, the coupler response of the completed hearing aids is compared to the random-incidence CORFIG curves which represented the design goals. The subsequently measured insertion gain is logically discussed at the same time, since it allowed a calculation of the 21-band accuracy score for the hearing aids. Pleasantly enough, the 80 percent goal was exceeded by the OTE aids, and the 89 percent goal was exceeded by the ITE aids!

A. Microphones

The OTE aids contained experimental EA-type microphones (XD-1116) which had been assembled to provide an approximately flat pressure response from 100 Hz to 8 kHz when coupled with 10 mm of 1.5 mm diameter inlet tubing. The frequency response of these microphones--with and without added tubing--is shown in the upper graph of Figure 2-12. These microphones were compliantly mounted in commercial OTE hearing aid cases, with roughly 1 m of miniature two-wire-shielded cable connecting each microphone to its amplifier.

The ITE aids contained BT-1759 microphones (sometimes called the "salt shaker" microphone because of the appearance of the multiple inlet holes in the top cover) whose frequency response is shown in the lower graph in Figure 2-13. The experimental ITE hearing aids were "molded in place" on the KEMAR manikin using "Mortite" compound (the earphone-earmold combination described below was imbedded in the compound), and the microphones were mounted flush with the surface, which was approximately in the plane of the pinna.
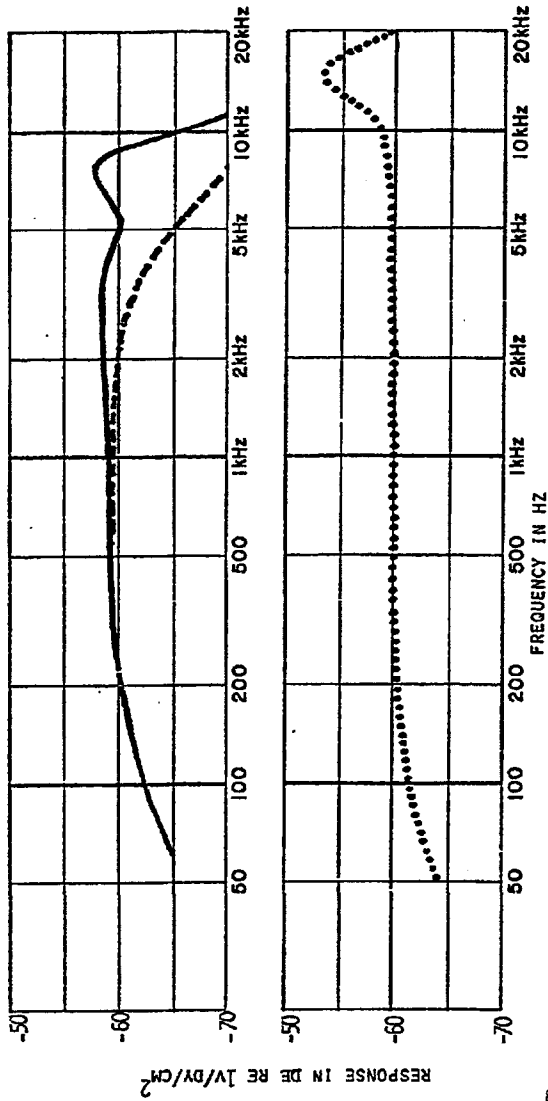
FIGURE 2-13 FREQUENCY RESPONSE OF MICROPHONES USED IN EXPERIMENTAL HEARING AIDS; XD-1116 WITH (———) AND WITHOUT (----) 10 MM OF 1.5 MM INLET TUBING, BT-1759 (·····).

RESPONSE IN DB RE [1v/DY/CM²]

85

B. Earphone-Earmold Combinations

For the OTE aids, BP-1712 earphones were compliantly mounted in the commercial OTE cases. A 10 mm length of 1.1 mm diameter rubber tubing coupled the earphone to the earhook. The earhook itself was modified to provide approximately 1.2 mm internal diameter over its 23 mm length. When coupled with an "8CR" earmold, the mounted earphone response shown previously in Figure 2-5 was obtained. (The construction of the 8CR earmold was shown inset in Figure 2-5.)

For the ITE aids, a matched pair of experimental low-impedance BP-type earphones was selected on the basis of their reasonably smooth high-frequency (above 8 kHz) response when coupled with a "16KM" earmold. Their frequency response was shown earlier in Figure 2-6. (The construction of the 16KM earmold was shown inset in Figure 2-6).

The low-impedance BP-type earphones used with the ITE aids had a 150 ohm nominal electrical impedance and were adjusted for 5 mA dc bias. This unusual combination was chosen to guarantee that amplifier overload would not occur on music peaks, and is not suggested for ITE design. The .7 mA dc bias used with the BP-1712 earphones in the OTE aids is entirely practical, however. (By way of reference, a .7 mA battery drain corresponds to nearly two weeks of 16-hour-per-day operation using a 1.5 Volt S76 battery).

Flexible shielded cable was used to connect each earphone to its associated external amplifier.

C. Amplifiers

The same basic 1.5V amplifier design was used with all of the hearing aids. The circuit configuration of these amplifiers is shown in Figure 2-14. The output amplifier is a high-gain inverting amplifier compensated (C2 and R4 in Figure 4a) to provide approximately 50 dB of stable feedback in the closed-loop condition. This unusually high degree of feedback was chosen to provide a low amplifier output impedance and to insure that the only limiting factor on overall distortion performance would be earphone linearity as long as operating levels were maintained below amplifier clipping.

The dc voltage maintained at the output terminal of the amplifiers is roughly .5V, so that the value of R7 can be chosen to produce the proper dc bias current in the earphone. The presence of R7 in series with the earphone also serves to provide a reduced low frequency response (for low-impedance drive sources), which was desirable in the present application. At the same time, the value chosen for R7 will generally be low enough that a low-impedance drive will be presented to the earphone: As discussed previously, the use of low-impedance drive to the earphone results in lower earphone distortion than could be obtained with high-impedance drive.

The preamplifier circuit (shown as a block in Figure 2-14) was set to produce a fixed 5 dB gain during the fidelity rating experiments. In order to produce the unity insertion gain (on the KEMAR manikin in a diffuse sound field) required for the experiments, the value of feedback resistor R3 was individually selected for each hearing aid to accommodate differences in transducer sensitivities.
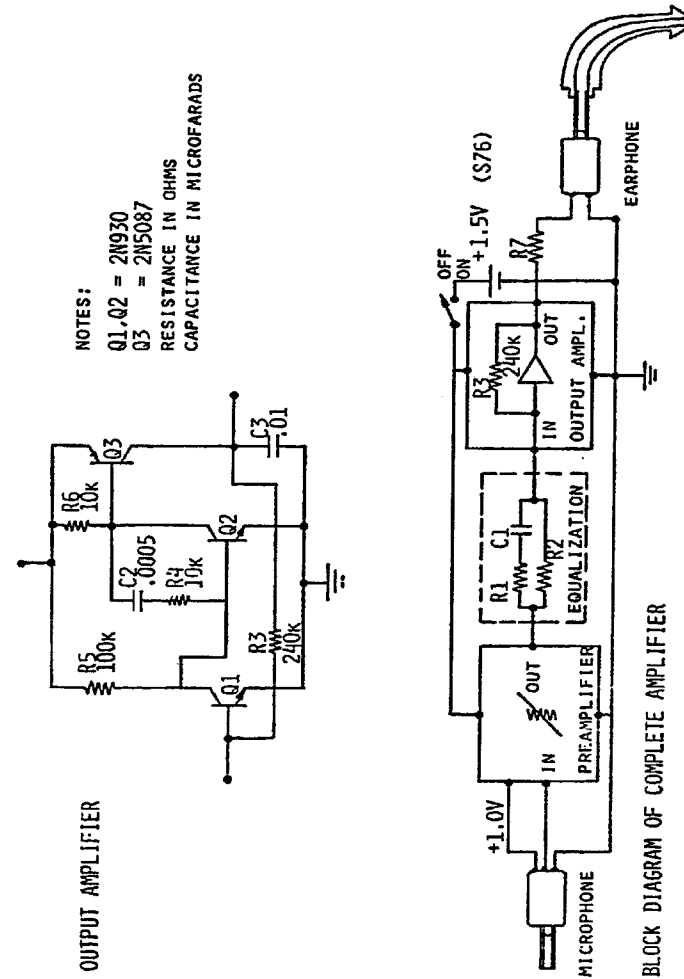
FIGURE 2-14   AMPLIFIER CONFIGURATION USED IN EXPERIMENTAL OTE AND ITE HEARING AIDS.

A simple equalization network (R1, C1, and R2 in Figure 2-14) provides the high-frequency response boost required to produce an approximately flat insertion-gain curve for the completed hearing aids. Table 1 shows the different circuit values used for the OTE and the ITE hearing aids, and Figure 2-15 shows the overall amplifier response with the OTE and ITE values. As would be expected, less amplifier (voltage) gain was required to produce unity insertion gain on KEMAR with the low impedance (150 ohm) earphones used in the ITE aids than with the moderate-impedance (1200 ohm) BP-1712 earphones used in the OTE hearing aids.

TABLE I

CIRCUIT VALUES FOR OTE AND ITE AIDS

	R1	C1	R2	R7	Earphone Bias
OTE	10K	.001	27K	330	.7 mA
ITE	12K	.001	82K	68	5 mA

D. An ITC Aid

An additional pair of "In-The-Concha-microphone-pickup" (ITC) hearing aids were included in the fidelity-rating experiments. These were similar to the ITE aids except the microphones were located in the bottom of the concha adjacent to the blocked (with a "canal" ear-mold) earcanal entrance and an external bridged-T filter was added
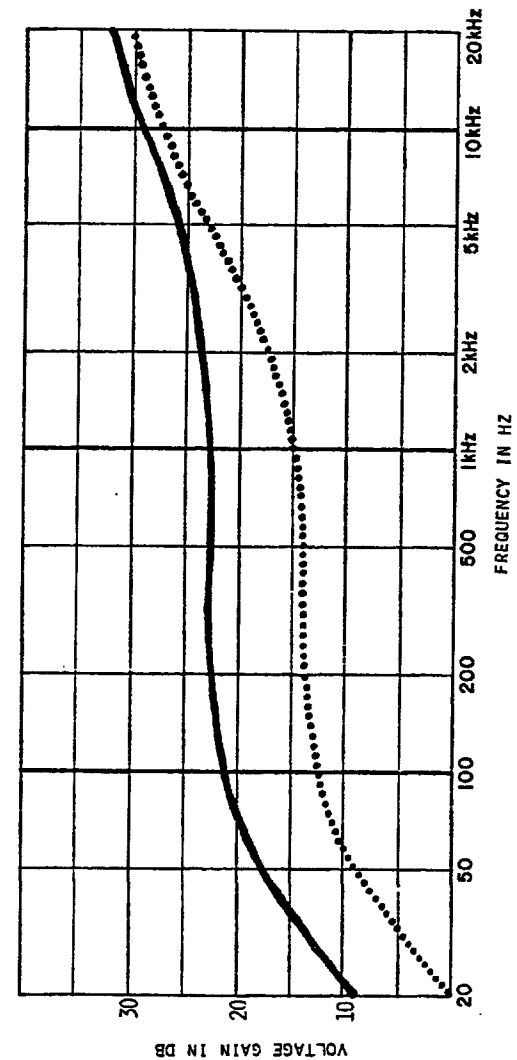
FIGURE 2-15 NOMINAL FREQUENCY RESPONSE OF COMPLETED OTE (———) AND ITE (••••) AMPLIFIERS

to equalize the 5 kHz peak introduced by the principal concha resonance. Since the In-The-Concha microphone location requires a nearly unoccluded concha, an ITE construction (including battery, amplifier, and volume control) would not be practical using today's components.

E. Frequency Responses of Completed Aids

The Zwislocki-coupler response of the completed OTE aids is shown in Figure 2-16 as a solid curve, compared to the random-incidence CORFIG response goal (dotted curve). The two agree within ± 3 dB up to nearly 8 kHz, which was the design cutoff frequency.

The Zwislocki-coupler response of the completed ITE aids is shown in Figure 2-17 (solid curve) compared to the random-incidence CORFIG response goal for ITE aids. Here it is clear that the simple amplifier compensation used with the ITE aids did not adequately compensate for the loss of external-ear resonance. (Simple equalization was adequate with the OTE aids because the compensation was designed into the 8CR earmold response characteristics).

Figure 2-18 shows the average insertion gain, measured in one-third-octave bands with a KEMAR manikin, of the OTE, ITE, and ITC hearing aids. The measurements were performed during the course of the listening-test recording sessions described below, and provide an estimate of the insertion gain a user would experience listening to a live concert performance or to a stereo high-fidelity system at home.

The calculated 21-band accuracy score corresponding to the OTE response curve of Figure 2-18 is 82 percent, 2 percent better than the "reasonable design goal" suggested earlier in order to match the
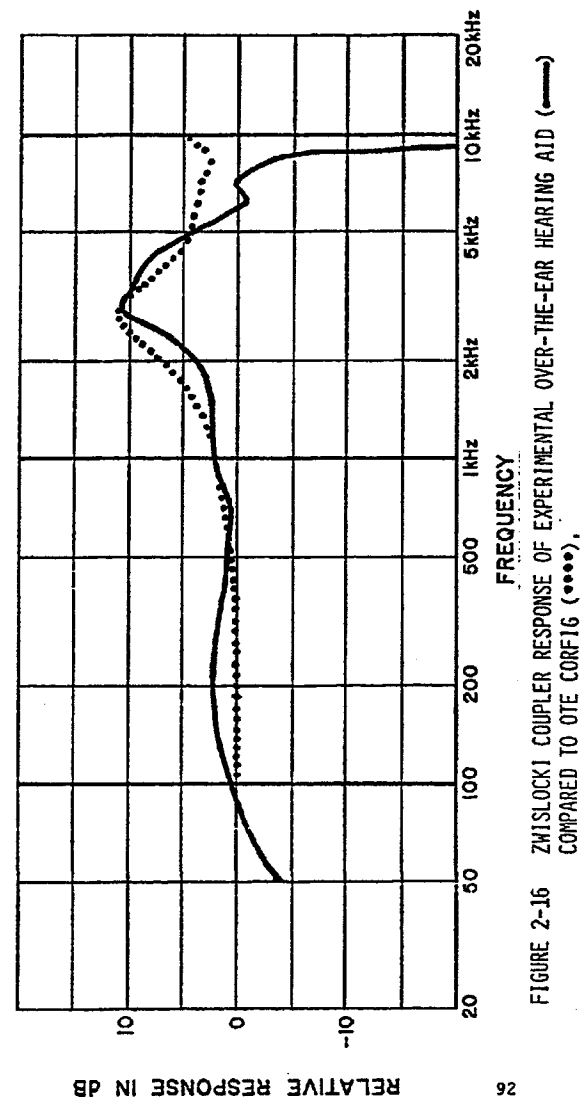
FIGURE 2-16 ZWISLOCKI COUPLER RESPONSE OF EXPERIMENTAL OVER-THE-EAR HEARING AID (——) COMPARED TO OTE CORFIG (•••).
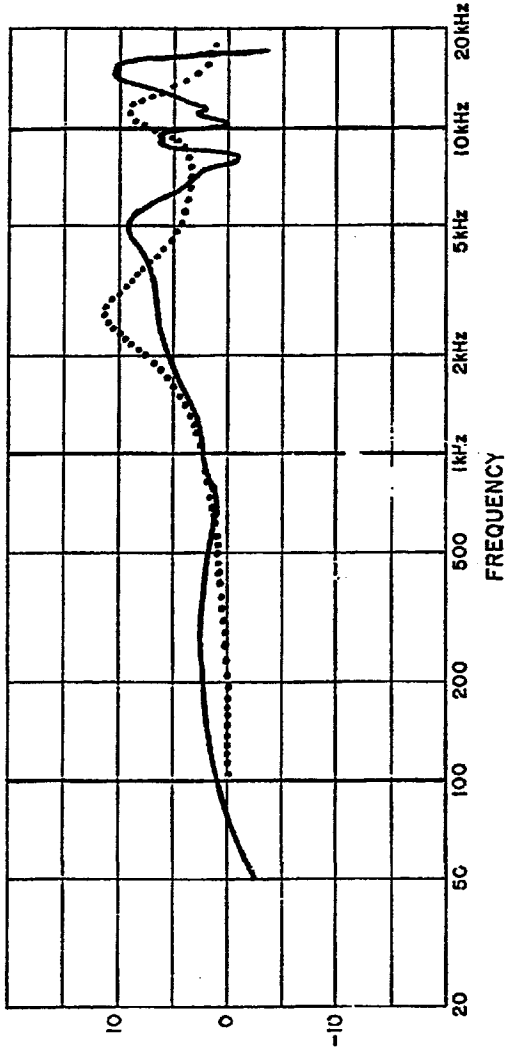
FIGURE 2-17   ZWISLOCKI COUPLER RESPONSE OF EXPERIMENTAL IN-THE-EAR HEARING AID (———)
COMPARED TO ITE CONFIG (••••).

RELATIVE RESPONSE IN dB
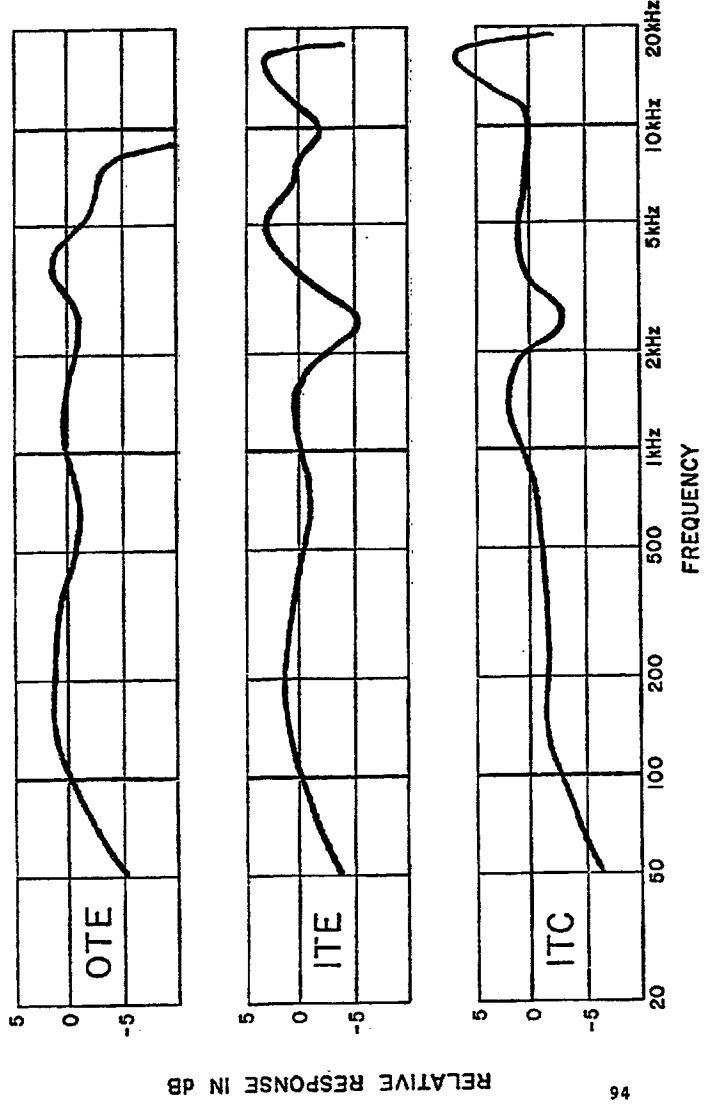
FIGURE 2-18   INSERTION GAIN OF EXPERIMENTAL HEARING AIDS.

NOTE: MEASURED ON KEMAR MANIKIN WITH ONE-THIRD-OCTAVE BANDS OF NOISE IN 170 M³
RECORDING ROOM.

RELATIVE RESPONSE IN dB

21-band accuracy scores obtained by the "inexpensive" loudspeakers.

The calculated 21-band accuracy score corresponding to the ITE response curve of Figure 2-18 is 91 percent, 2 percent better than the "stringent design goal" required in order to match the 21-band accuracy score of the median state-of-the-art "expensive" loudspeakers. The principal inaccuracy evident in the ITE insertion-gain curve is the dip at 2.7 kHz, the result of the inadequate amplifier equalization discussed above.

The use of more sophisticated amplifier compensation, in this case a bridged-T filter, along with an In-The-Concha microphone location can result in a more accurate insertion gain, as shown by the ITC curve in Figure 2-18. Although the calculated 21-band accuracy score (92 percent) improved by only one percent compared to the ITE curve, this was due to a failure to correct a 1 dB level error in the 150- to 800-Hz region. That correction would have provided another two or three percent improvement. (The simple amplifier equalization required with the ITE aid could easily be included in a practical In-The-Ear hearing aid amplifier. The In-The-Concha microphone location requires a nearly unoccluded concha, however, which would rule out an "In-The-Ear" construction for the ITC aid using today's components. Such a construction cannot be ruled out for future designs, however.)

F. Distortion of Completed Aids

Because of the large amount of negative feedback and low output impedance used in the experimental amplifiers, distortion was not expected to be a problem in the experimental hearing aids. (Recall that

the distortion of the BP-series earphones themselves is low compared to the ear, as illustrated in Figure 2-11, p. 72). Nonetheless, the nonlinear distortion of one of the OTE aids was examined in some detail as a check.

Plots of second- and third-harmonic distortion versus frequency were obtained for input sound-pressure levels of 60, 70, 80, 90, 100, 105, and 110 dB. At no frequency did hearing aid distortion--measured in a Zwislocki coupler--exceed one percent for inputs of 100 dB SPL or less. In most cases, the distortion levels were close to the measurement limits imposed by the B&K 2020 Heterodyne Analyzer and 1902 Distortion Control Unit used to obtain automatic distortion-level plots. Plots of CCIF-intermodulation distortion obtained for a 200 Hz difference frequency showed a similar result: Below 100 dB SPL, the distortion typically fell below the measurement limits of the B&K equipment.

Data on total harmonic distortion versus output--measured in a Zwislocki coupler--were also obtained at a fixed 500 Hz input frequency. Those data are shown plotted in Figure 2-19. Below 105 dB SPL output, the measured total harmonic distortion is roughly one-fourth the estimated maximum inaudible (for music and speech) hearing-aid distortion levels shown earlier in Figure 2-20 and reproduced as a dashed curve in Figure 2-19. The abrupt increase in measured distortion above 105-dB SPL corresponds to the onset of amplifier clipping. As noted earlier, the BP-series earphone itself is capable of substantially higher undistorted output (see, for example, Figure 2-8, p. 63). The curve shown in Figure 2-19 reflects the writer's choice of earphone

impedance and battery drain. The .7 mA battery drain used with the
OTE aids was sufficient to meet the undistorted 105-dB instantaneous-
peak SPL design goal. Note that the two-percent distortion which
occurs at 105 dB sine-wave SPL in Figure 2-19 corresponds to an instan-
taneous peak 3 dB higher, or 108 dB SPL. Greater undistorted output
is possible with increased battery drain and/or the use of a Class B
output amplifier.

In any case, all the design requirements for high fidelity hearing
aids had been met by the experimental aids. On paper, at least, they
would qualify as high fidelity by any reasonable standards. The only
remaining question was whether or not they would sound like a high-
fidelity sound reproduction system. Although extensive listening tests
at Ravinia and elsewhere on these and earlier (less accurate) versions
of these aids had convinced the writer, a more formal evaluation was
clearly called for. A listening-test experiment designed to provide
such an evaluation is discussed in the following chapters.

FIGURE 2-19 TOTAL HARMONIC DISTORTION AT 500 HZ VERSUS OUTPUT FOR THE OTE HEARING AIDS (———) COMPARED TO ESTIMATED MAXIMUM INAUDIBLE DISTORTION (– – –).

CHAPTER III

FIDELITY RATING -- EXPERIMENTAL DESIGN

Recall that the primary thrust of this investigation was toward a demonstration that transducer and amplifier technology had advanced far enough that it was now possible to build hearing aids which would be classified as high-fidelity sound reproducers by those with normal hearing. In addition to the objective data presented in the preceding chapter, such a demonstration requires some form of subjective fidelity rating experiment.

In this chapter, several experimental design considerations relating to the experimental validity and statistical reliability of such a rating experiment are discussed. The three sections in this chapter are titled: Experimental Approach, Questions and Hypotheses, and Statistical Considerations.

I. EXPERIMENTAL APPROACH

A. Prerecorded Comparisons

The technique with the greatest face validity for rating the fidelity of a sound reproduction system is to compare the reproduced sound with the original sound. This approach was used by Olson (1957) in his famous 1947 demonstration, in which the Boston Symphony Orchestra was compared with a phonograph record of the orchestra before an overflow audience in the music shed at Tanglewood, Massachusetts. This live-versus-recorded comparison technique was subsequently used

extensively by Villchur (1964) in the design and demonstration of the AR3 loudspeaker system.

In this context, it is important to keep in mind the distinction between sound quality and sound fidelity. One may or may not like the sound quality achieved by the Boston Symphony Orchestra or the Fine Arts Quartet--live or recorded--but a perfect fidelity reproduction should be indistinguishable from the original.

While a true live-versus-recorded listening test has excellent face validity, it becomes impractically cumbersome when several different sound reproduction systems are to be tested. Each subject would have to be supplied with a sample of each pair of hearing aids or earphones to be tested, for example, while the large number of repeated comparisons dictated by multiple systems and multiple program selections would tax the patience of any musical group. A more tractable listening-test experiment results from the use of tape recordings, both for the source material and the comparisons themselves.

Villchur (1962) used pre-recorded stimuli in a "simulated live-versus-recorded" technique, where the source material was itself a reproduction of previously recorded material. To use this technique for comparisons employing musical reproductions, for example, a loudspeaker with good dispersion (output nearly the same in all directions) is chosen as a "reference" loudspeaker. Anechoic chamber recordings of that loudspeaker reproducing musical selections from a Master Tape are then obtained, just as if that reference loudspeaker were itself a group of live musicians. The simulated-live-versus-recorded comparisons are subsequently presented between (a) the reference loudspeaker

reproducing the original Master Tape recording (the simulated live source) and (b) the loudspeaker under test reproducing the anechoic-chamber recording of that simulated live source. Note that the exact frequency response of the reference loudspeaker is not important (as long as its bandwidth is sufficient to exercise the loudspeaker system under test), just as it would be immaterial to a true live-versus-recorded comparison whether the violinist chooses a violin made by Stradivari or Guarneri. (The reason good dispersion is desirable is that it simplifies the task of finding the anechoic-chamber microphone location which best represents the total power output of the reference loudspeaker.)

Villchur's simulated-live-versus-recorded technique removes the necessity for having live musicians present for each comparison presentation. A further simplification can be obtained if the loudspeaker system chosen for the "surrogate live source" is known to have a sensibly flat frequency response, in which case the anechoic-chamber recording may be dispensed with. That approach was chosen for the present experiments, recognizing that the validity of the resulting comparisons depends heavily on the accuracy of the loudspeaker system chosen for the surrogate live source if they are to be interpreted as fidelity ratings. That requirement is discussed in more detail below.

When stereo-loudspeaker reproduction provides the "surrogate live source" and a KEMAR manikin is used as a "surrogate subject", the result might be labeled "manikin-prerecorded-simulated-live-versus-recorded" listening tests. The acronym MAPSLIVER is clearly available, but such tests will be described below simply as "prerecorded comparisons."

1. Requirements of surrogate live source

Assuming the manikin is a reasonable representation of average human acoustics, comparison between recordings obtained from eardrum-position microphones in the manikin with and without hearing aids in place allows direct rating of the fidelity of the hearing aids as sound reproducers. Little demand is placed on the accuracy of the surrogate live source for such aided-unaided comparisons, since the same source is used in both cases. When comparisons to other loud-speakers are intended, however, the "reference loudspeakers" used for the surrogate live source must have good fidelity if the ratings are to have meaning as "fidelity ratings" in the usual sense. Little demand is placed on the room acoustics for such comparisons, however, since the same room affects all loudspeakers.

If comparison to headphone-reproduction is also desired, then not only must the reference loudspeakers have good fidelity per se, but the acoustics of the room must be reasonable; i.e., the loudspeaker-room system must have good fidelity.

Finally, the surrogate source must be capable of reproducing the program material at original (i.e. "live") levels without audible distortion.

The obvious choice for the reference loudspeaker system would be one which itself had successfully passed true live-versus-recorded listening tests. As mentioned above, the AR3 and AR3a loudspeakers fall in this category.

A well-designed professional recording studio would appear to be the ideal "listening room" for several reasons: a) it will have

reverberation time (as a function of frequency) optimized for its volume; b) it will be free of "flutter echoes"; and c) partly as a direct result of a) and b), it will have a uniform "room response". With a choice of 6000 cubic feet for the studio volume, the optimum reverberation time (Olson, 1967; p. 307) will be equal to the average .3 to .5 seconds found in typical living rooms (Moir, 1958; p. 509).

2. Requirements of surrogate listener

Although pre-recorded comparisons among hearing aids have often been employed in the past, only recently has it been possible to employ a suitable manikin for recording the hearing aid output. The success of the KEMAR manikin in duplicating average human acoustics has been well documented. Most recently, Cox and Studebaker (1979) studied the similarity in spectra between signals processed by a hearing aid in a "live" situation and signals first tape recorded with the same hearing aid worn by a KEMAR manikin and then filtered and reproduced through a hearing aid earphone with custom earmold. They found essentially similar results could be obtained as long as well-sealed earmolds were used.

For our present purposes, the pre-recorded comparisons may be reproduced over loudspeakers or headphones. Any earmold-earphone-ear interactions would thus not be duplicated. Stated somewhat differently, a hearing aid which had perfect fidelity for the hypothetical average user (or the KEMAR manikin) would not have perfect fidelity for every user because of individual differences as discussed in Chapter II, Section II-D. No investigation of the importance of these differences was planned during the present experimental design.

Another potential source of error in the present experiments is that the modified Zwislocki coupler (IRPI DB 100) is only known to be representative of real ears up to approximately 8 kHz. The suitability of that coupler in the octave band from 8 to 18 kHz is unknown at present. Although the coupler accuracy would have little impact on loud-speaker-loudspeaker comparisons, it would affect any earphone-versus-loudspeaker or hearing-aid-versus-loudspeaker comparisons.

3. Requirements of playback reproducers

Subjective fidelity ratings can be obtained from subjects listening to a playback of the prerecorded comparisons over either loud-speakers or earphones.

If the prerecorded comparisons are presented as AB or ABA comparisons, then the fidelity-rating task is equivalent to a similarity-judgment task (as it would be in a true live-versus-recorded experiment). Under those circumstances, the choice of playback reproducers should be much less critical than the choice of the reference loudspeaker system; any coloration introduced by frequency-response inaccuracies would be expected to affect both the sound of the reference and the comparison systems about equally. The most important requirements are that the distortion and response irregularities of the playback reproducers (or playback-room acoustics) be low enough to avoid the possibility of masking any distortion produced by one of the comparison sound systems, and that the effective bandwidth of the reproducers include all frequencies of interest in the comparisons (in this case the full audio band from perhaps 40 Hz to 14,000 Hz).

When the comparison recordings are true binaural recordings obtained from eardrum-position microphones in a realistic manikin, they should ideally be reproduced by earphones: earphones having a flat eardrum-pressure frequency response. In order to conveniently obtain ratings from large groups of subjects, however, loudspeaker reproduction is indicated.

A problem arises with either type of reproducer, since commonly available loudspeakers and headphones are designed to produce a relatively flat frequency response referred to the sound field. As a result, their eardrum-pressure frequency response will exhibit a peak of roughly 15 dB at 2.7 kHz due to the effect of external-ear resonances (Shaw, 1976). When added to the roughly 15 dB peak introduced by the external-ear resonances in the manikin, a duplication of resonances occurs. (The subjective result of such a duplication is a single 15 dB peak, since the peak introduced by the subjects' own external-ear resonances is a normal part of his listening experience.)

Although the same peak would be added to both the reference and comparison sounds, such a large peak is likely to introduce a bias in favor of systems with compensating deficiencies. Moreover, equalization of the manikin recordings is required if the reference system is to sound plausible as a surrogate live source. These problems were discussed by Killion (1979a). In that paper, a simple bridged-T filter was shown to provide equalization accurate to ±3 dB in a diffuse sound field; an accuracy adequate for most listening-test purposes.

The use of loudspeaker playback instead of headphone playback of the prerecorded comparisons can be expected to weaken the sensitivity of the listening tests to the loss of "listening-room ambience" occurring with manikin-headphone recordings. With loudspeaker playback, the presence of reverberation, early reflections, etc. in the room used for presentation of the listening-test comparisons will tend to cover up the lack (in the manikin-headphone recordings) of "listening-room" (recording studio) reverberation, reflections, etc.

Thus loudspeaker-hearing-aid and loudspeaker-loudspeaker comparisons would be expected to produce relatively higher ratings vis-a-vis headphone-loudspeaker comparisons when the comparisons are presented over headphones. Whether this is good or bad depends entirely on one's point of view: the lack of listening-room ambience found in headphone listening is considered a disadvantage by some because it makes for an "unnatural sound", while others consider it an advantage because only the original concert-hall ambience is reproduced. Either type of listening-test playback thus appears defensible.

A final consideration is the level chosen for the original comparison recordings and the subsequent listening-test playback. For greatest face validity in the fidelity ratings, both should come as close as possible to the original levels that a typical listener would experience listening in a "live" situation. Moreover, even though Olson (1957) reported that live-performance levels are typically 20 dB greater than the levels people choose for home listening, Gabrielsson and Sjogren (1976) found that reduced-level listening tests were less revealing of system deficiencies than original-level tests.

4. Requirements of tape recorders

The creation of the final AB or ABA comparisons is simplified by the use of a four-track recorder to record the output from the manikin. With this approach, the reference loudspeaker reproduction can be recorded on one pair of tracks, the tape can be rewound, and the comparison reproduction can be subsequently recorded in synchrony on the other pair of tracks. The final comparisons can then be easily recorded on a separate two-track recorder by simply switching from one pair of tracks to the other at the appropriate time(s) during playback of the four-track recording.

Because of the multiple generations of tape recording required by the prerecorded listening tests, extreme care in maintaining the alignment and equalization of the tape recorders is required. The matching of the pairs of tracks on the four-track recorder is particularly important; audible differences between track pairs would tend to vitiate the comparisons.

As mentioned above, distortion in the recording or playback system could mask distortion produced by one of the comparison systems. A prime candidate for such distortion is the tape recording process, as discussed by Burnett, Corliss, and Berendt (1972). Simply insuring that program peaks do not exceed "0 VU" does not guarantee distortion-free recordings with high-peak-factor sources such as the piano, or when a non-uniform frequency response produces substantial high-frequency emphasis. The high-frequency pre-emphasis built into tape recorders uses up much of the "headroom" above 0 VU at high frequencies, particularly with a slow tape speed such as 7.5 inches per second.

In order to reduce the probability of tape overload to negligible levels, a 15 ips tape speed and extremely conservative recording levels (keeping peaks below -7 dB VU) were suggested by Burnett et al. With the greatly increased saturation output capabilities of more recent mastering tapes, the same result can be obtained by simply calibrating the 0 VU setting on the recorder conservatively. A 0 VU calibration at 200 nW/m is approximately 15 dB below the three percent total-harmonic-distortion level (and 20 dB below the saturation level) of Ampex 456 Grandmaster tape, for example.

Fortunately, the use of modern mastering tapes and professional tape recorders such as the Ampex 440 series of recorders allows the use of conservative recording levels without the tape noise becoming a serious problem. The A-weighted noise level of the Ampex 456 Grandmaster tape is some 60 dB below a 200 nW/m reference level. If the 0 VU recording level is made equivalent to a 94 dB SPL in the sound field, (a level which allows a typical fully-orchestrated fortissimo passages of 95 dB SPL to hit +1 dB VU), the tape noise level will be equivalent to a 34 dB(A) SPL referred to the sound field. Even after the second-generation comparison recordings are made, the tape noise level should be equivalent to no more than a 37 dB(A) SPL. That level is comparable to the 32 dB(A) noise level found in typical concert halls, and is thus low enough to preclude the masking of important distortion products, yet high enough to itself mask the roughly 25 dB(A) SPL equivalent noise level of the hearing aids. The latter masking is desirable in listening tests in order that differences in system bandwidth not be exposed by the wideband background noise

rather than the program material itself (Bauer, 1945).

B. Selection of Comparison Systems

The phrase "high fidelity" has no black-or-white definition; in practice, fidelity comes in various shades of grey. Furthermore, different systems are likely to have different imperfections. Although one could easily imagine a percentage-fidelity rating scale, an isolated fidelity rating of a hearing aid would be of little interest to anyone. Some idea of how other commonly experienced high-fidelity systems are rated by the same test is required in order to provide a) a working definition of the fidelity rating and b) a "calibration" of the test itself. Both requirements can be satisfied by including a range of popular "high fidelity" systems in the fidelity-rating experiment to serve as benchmarks against which to compare the fidelity rating given the experimental hearing aids.

In order to provide a set of benchmark systems, it seems reasonable to include the following sound systems in the main experiment (in addition to the OTE and ITE hearing aids):

1. A five-dollar pocket ratio as a within-scale low-fidelity anchor to help define the meaning of the ratings (Guilford, 1954);

2. An inexpensive (under $100) "high fidelity" stereo phonograph;

3. A speech audiometer assembled from a professional tape recorder, amplifier, and a carefully selected pair of TDH-39 headphones in MX-41/AR cushions;

4. A popular Koss stereo headphone with known (designed-in) exaggeration of the bass response;

5. An expensive ($500-1000/pair) monitor-loudspeaker system with frequency response known to be different from that of the reference loudspeaker system.

C. Selection of Program Material

Some variety in program material is required simply in order to provide some representation of the material to which people commonly listen. More importantly, no one type of program material can adequately expose all system deficiencies.

The use of live voice provides a very practical fidelity test for both hearing aids and loudspeakers, but has an audible bandwidth of only 100 to 8000 Hz (Snow, 1931). The use of a jazz piano trio provides a source which makes comparison of the bandwidth of various systems relatively easy: The string bass produces audible energy below 100 Hz, while the brush-on-cymbal sound of the drummer produces audible energy above 8 kHz. A fully scored orchestral passage provides a source which facilitates comparison of the smoothness of the midband frequency response of various systems: A fully scored orchestral passage is relatively dense in the frequency domain. Wideband noise--which is the densest in the frequency domain--provides a source for listening tests which are extremely sensitive to the coloration produced by peaks and dips in the frequency-response curve (Villchur, 1962), but wideband noise is not the sort of material to which people commonly choose to listen.

One sound system may have a wide bandwidth but a ragged frequency response, or exaggerated bass and treble response (typical of some of the better-selling "high-fidelity" reproducers). Such a system could be expected to rate high on piano-trio reproduction but lower on orchestral or wideband noise reproduction. Another sound system may have a smooth response but limited bandwidth, a limitation which tends to be more evident on the piano trio than on live voice or even the full orchestra. Still another system may have smooth response and a good high-frequency bandwidth but weak bass response. Subjects who "like a lot of bass" are likely to rate that system lower than subjects who concentrate on the high-frequency response.

The four different types of program material discussed above were chosen for the present experiment.

D. Selection of Subjects

One could argue equally well for the use of two entirely different types of subjects. A listening-test jury comprised of randomly selected subjects would presumably produce ratings with the widest applicability since they would represent the "man-on-the-streets" judgment.

A listening-test jury comprised of so-called "golden ear" subjects --listeners highly trained at the subjective evaluation of high-fidelity systems--would presumably produce ratings which more nearly reflected the "true" fidelity (on some imagined absolute scale) of the rated systems.

The training of the golden-ear subjects involves at least two different facets; learning to retain a good auditory image of live performances, and learning to detect and evaluate the importance of a wide variety of potential defects in sound-reproduction systems. Gabrielsson, Rosenberg and Sjogren (1974), however, reported no significant difference on the average over all listening conditions between "high-fidelity" subjects and "listeners in general" in a five-loud-speaker fidelity-rating experiment utilizing five musical selections. In that experiment, the reference for these ratings was the subject's memory of how a live performance sounded based on past experience.

In a later fidelity-rating experiment, Gabrielsson and Sjogren (1976) reported significant but relatively small differences between trained and untrained listeners in terms of the overall average ratings, although the trained-listener's judgments were much more reliable (i.e., a smaller number of trained listeners was required to establish a fidelity rating within a given confidence interval).

Given these results, the greater applicability of man-on-the-street ratings argues for the use of as close as possible to a random sample of the normal-hearing population at large, providing only that it is possible to employ enough subjects to produce sufficiently reliable ratings.

E. Fidelity Rating Scale

A wide variety of scales have been used in fidelity-rating experiments, but the most common is probably some variant of the simple zero to 100 percent scale. Snow (1931) presented the results of his

acceptable-quality-versus-bandwidth experiments using such a scale. (The original ratings were obtained using a zero to 1.0 scale, with provision for use of numbers greater than 1.0 if the listeners felt the bandwidth restriction represented an improvement, but ratings exceeding 1.0 were rare).

Gabrielsson, Rosenberg, and Sjogren (1974) used a scale of zero to ten (with decimal ratings allowed) in the fidelity-rating experiment mentioned above. In that experiment, only the endpoints of the scale were defined, with zero corresponding to "totally unrecognizable reproduction" and ten corresponding to "perfectly true-to-nature reproduction". Using this scale, the average ratings given five loudspeakers ranged from 3.6 for a "poor" fidelity loudspeaker through 5.6 and 5.9 for two "medium" fidelity loudspeakers to 7.4 and 7.5 for two "high" fidelity loudspeakers. (The subject's memory of how a live performance sounded provided the reference for these fidelity ratings.)

In a subsequent experiment, Gabrielsson and Sjogren (1976) provided adjective definitions at several points along the scale as follows: 9--Excellent; 7--Good; 5--Fair; 3--Bad; 1--Very bad. They also changed the definition of zero to "practically no similarity to live performance," a change which served to allow greater use of the lower end of the scale with common reproducers. (In today's world, a reproducer which provided totally unrecognizable performance might be hard to find).

The adjective-defined rating scale of Gabrielsson and Sjogren appears suitable for the present purposes. In order to provide a rating scale whose appearance is more familiar to untrained listeners,

however, it seems advisable to use a true zero to 100 percent scale. (Since Gabrielsson et al. allowed decimal ratings, a whole-number percentage scale would be numerically equivalent to their rating scale multiplied by the factor ten).

II. QUESTIONS AND HYPOTHESES

A list of experimental questions which might reasonably be asked of a fidelity rating experiment is given in this section, along with the traditional statement of the corresponding null hypotheses to be tested.

A. A List of Questions

The following questions were considered important:

1. Can the experimental hearing aid(s) be considered a high fidelity system in the sense that it receives untrained-listener fidelity ratings in the listening test comparable to those given other sound systems commonly labeled "high fidelity"?

2. Does the same conclusion hold for ratings obtained from trained listeners?

3. What effect does the use of different program materials have on the fidelity ratings?

4. Are there interactions between sound systems and program materials; systems and listeners; programs and listeners; or, systems and programs and listeners?

5. What is the reliability of the fidelity ratings obtainable in a subjective experiment of this sort?

6. What is the validity of the ratings; do they appear likely to
represent--within reasonable bounds--the ratings which an
average listener might give in an unstructured real-life
listening situation?

All except the last question can be answered on the basis of a
statistical analysis of the fidelity rating experiment, and can be re-
stated in terms of null hypotheses. The last (validity) question
requires a judgmental answer which each reader must supply for himself
after reviewing the experimental design and execution.

B. The Null Hypotheses

The following hypotheses are readily tested with the aid of an
analysis of variance applied to the listening-test results:

H_1: There are no differences among the ratings of the various
sound systems.

H_2: There are no differences among the ratings obtained with
different program materials.

H_3: There are no differences among the ratings produced by
different subjects.

H_4: There are no interactions between:

a) sound systems and program materials

b) systems and subjects

c) programs and subjects

d) systems, programs, and subjects

Answering the main question requires a careful choice of benchmark
systems to be included in the listening test. Since the speech audiometer

has often been used as a standard of comparison for fidelity in the
audiological literature, a comparison of the hearing aids to a nearly
flawless speech audiometer might be considered the most important com-
parison in an audiological context. The null hypothesis corresponding
to such a comparison can be stated as follows:

H_5: The fidelity rating given either of the two experimental
hearing aid pairs (OTE or ITE) is not different than that
given the speech audiometer.

Rejection of Hypotheses H_5--in combination with a greater rating
for at least one of the hearing aids--would allow the conclusion that
it was possible to design hearing aids which at least exceeded the
standard of fidelity set by a speech audiometer.

A more general answer derives from an estimation of the fidelity
of the hearing aids relative to more conventional "high-fidelity" sys-
tems. If it is granted that it may be possible to design hearing aids
with a fidelity better than something which has been labeled high
fidelity, the real question becomes where the hearing aid rating falls
on the fidelity continuum. In this case, the immediate problem is one
of estimation, not a direct test of hypotheses. As stated by Davies
et al. (1960):

> (When) it is known. . .that some variation must exist, there
> would be no point in testing the hypothesis that it did not
> exist. . .(moreover,) it will be found. . .that even when it
> is reasonable to test a Null Hypothesis it may be better to
> calculate confidence intervals and draw conclusions from
> these. (p. 115)

Since the confidence intervals derive from the same statistical
analysis used for testing hypotheses (Brownlee, 1965), there is

probably little harm in stating the more general fidelity-placement questions in terms of a final null hypothesis, however:

H_6: The fidelity rating given either of the two hearing aids is not different than that given any of the other comparison high-fidelity systems.

III. STATISTICAL CONSIDERATIONS

A. Confidence Levels

The first statistical decision which must be made is the confidence level to be used in the test of hypotheses. For the present experiment, a confidence level of 95 percent (.05 significance level) was chosen for all tests except that of Hypothesis H_5, where a confidence level of 99 percent was chosen. This more stringent confidence level for hypothesis H_5 was chosen for two reasons. First, the writer wished to particularly avoid an alpha-type error (in this case, incorrectly accepting the "alternate hypothesis" that the fidelity rating for either hearing aid was actually greater than that for the speech audiometer), and was relatively unconcerned about a possible beta-type error (in this case, failing to find a significant difference when one actually existed). The second reason was one of convenience. Only two comparisons were required for the test of hypothesis H_5, whereas ten comparisons were required for a test of hypothesis H_6. As discussed below, the same critical difference can be used to test two hypotheses at the 99-percent level as ten at the 95-percent level. /In truth, the latter was the main reason for the choice, a confession the writer feels emboldened to make after running across the following apology

in Miller (1966, p. 68): "Although scaling down the significance level when confronted with more than one statement is an old device, it is seldom taught to students and is rarely mentioned in textbooks and the literature."/

B. Sample Size

Preliminary listening-test experiments indicated the standard deviation across subjects on a single rating would be in the neighborhood of 15 percent for college-educated listeners. Assuming a more conservative 20 percent for an average listener, a 24-subject listening jury would allow determination of the mean "within-cell" rating (of any one system reproducing one program selection) to within ± 8 percent at the 95 percent confidence level. /The 95% confidence interval is $\pm 1.96 \, \sigma_{\bar{x}}$. With a sample size of 24, the standard error of the mean ($\sigma_{\bar{x}}$) would be $20\%/\sqrt{24} = 4.1\%$./

In order to compare two sound systems across all program materials (considered as a fixed effect for purposes of the statistical analysis), a t-test is normally applied with the standard error of the mean estimated from the subject-system interaction mean square obtained from a three-way analysis of variance (systems, programs, and subjects). Such an analysis applied to the preliminary listening test experiments indicated that the inclusion of six program selections might produce an interaction mean-square of 400 percent, corresponding to a standard error of the across-program system means of 1.7 percent with 24 subjects.

With 24 subjects and a total of seven comparison systems (two

hearing aids, four high-fidelity systems, and one low-fidelity system),

the t-value has 138 degrees of freedom associated with it, so it will

be nearly equal to the Normal-distribution value. Thus the critical

difference in the above example would be 1.96 $\sqrt{2}\ \sigma_{\bar{x}}$ or 4.7 percent at

the 95 percent confidence level. Since the expected difference--based

on the preliminary experiments--between the fidelity rating of the

speech audiometer and either hearing aid was 20 to 30 percent, a test

of hypothesis H_5 at the 10σ level (!) would presumably be possible

with a 24-subject x 6-program rating experiment.

C. Multiple Comparisons

A speech audiometer is not a true high-fidelity system, however,

so the multiple comparisons of hypothesis H_6 are required for the more

general answer to the main question. By virtue of the Bonferroni

inequality (Miller, 1966), the ten simultaneous comparisons required

by hypotheses H_5 and H_6 can be tested at the .05 significance level

(95% confidence level) using the .005 significance-level criterion for

a single comparison. (The probability that any one of ten comparisons

will meet a given criterion level is no greater than ten times the

probability for a single comparison. Thus, a .005 probability for a

single comparison becomes at most a .05 probability for ten.) Using

that approach with the above estimates, a t-test comparison between the

hearing aids and each of the other sound systems would allow rejection

of null hypothesis H_6 if a hearing aid rating differed from one of the

others by 2.8 $\sqrt{2}\ \sigma_{\bar{x}}$ or 6.7 percent in the above example.

It is interesting to note in passing that the normal probability

distribution approaches zero so rapidly in the tail regions that even

the $(6-1)! = 120$ possible two-way comparisons between the six "high-

fidelity" systems could be made at the 95 percent confidence level with

a critical difference of only 8.5 percent using the Bonferroni inequal-

ity in the above example.

It is also of interest to estimate the power of the above test for

differences of ten percent or more in the "true" ratings, a difference

slightly in excess of the eight percent which Consumers Union reported

was "difficult to resolve by ear" in their listening test evaluations

(Consumers Reports, October 1977). For a single comparison, the risk

of missing a true difference of ten percent (using a 6.7 percent cri-

terion for significance) would be approximately 15 percent. On the

other hand, the risk of missing a true difference as large as 13 percent

in any of the multiple comparisons would be less than five percent.

D. Confidence Limits

As observed above, the confidence limits for any of the mean rat-

ings derives directly from the same analysis used for testing hypothe-

ses. Thus the 95 percent confidence limits for any one of the six

across-programs mean system-fidelity ratings is obtained from the

$\pm.0025/6 = .0041$ significance-level points on the normal distribution.

In the above example, the true mean rating for each system would pre-

sumably be known within $\pm 2.64 \times 1.7$ or about 4.5 percent at the 95

percent confidence level.

On all accounts, a 24-subject, 6-program, and 7-system experimental design appeared adequate to the present purposes.

E. A Computer Program

The writer was unable to locate a three-way analysis of variance computer program which would readily accept raw data from a randomized-block experiment, sort it into the appropriate cells, allow ready choice of a subset of blocks to be analyzed, and then provide output in a convenient format. Appendix A contains the listing of a pair of Fortran Programs written for that purpose.

The "SORT" program produces a decoding file for the location of data in the randomized blocks. The "MANOVA" program accepts a complete data file, sorts the data according to the SORT file, and performs a three-way analysis of variance on the sorted data. An arcsin data transformation is an option. The program automatically collapses into a two-way analysis of variance when data from only one subject are entered or data from only one column (program selection) are selected.

CHAPTER IV

SUBJECTIVE FIDELITY-RATING EXPERIMENTS

I. METHOD

A. Stimuli: Master Stimulus Tape

Six selections of program material were spliced together to form a Master Stimulus Tape.[1] One selection was an anechoic chamber recording of repeated nonsense sentences ("Joe took father's shoebench out; she was sitting at my lawn.") spoken by the writer at a distance of 1m from a pair of Shure Brothers SM-81 cardiod condenser microphones arranged in the "ORTF" configuration (17 cm spacing, 110° included angle; Ceoen, 1972). One selection was 15 seconds of "speech spectrum noise" (broadband noise filtered to provide approximately the long-term average speech spectrum) recorded directly from the electrical output of a Grason-Stadler Model 901B noise generator.

The remaining four selections were musical passages which were dubbed from virgin pressings onto the Master Stimulus Tape by a professional recording studio using a Shure V15-4 cartridge. Two of the passages were taken from a N.Y. Philharmonic recording of the Beethoven Violin Concerto in D (Columbia stereo record M33587) and two from an Oscar Peterson piano trio recording of Peterson's blues "The Smudge" (Mercury stereo record EMS-2-405). One of the orchestral

---

[1] Ampex 456 "Grandmaster" tape, with a OVU recorder calibration of 200 nW/m, was used throughout.

passages was fortissimo, the other was forte. An attempt was made to select passages (orchestra and piano trio) which would be relatively unchanging through the switchover region from the reference system A to the comparison system B in order to allow the most sensitive A-B comparisons. (A recording containing comparisons obtained with the OTE hearing aids reproducing each of the six program selections can be obtained from the writer by anyone wishing to hear these six selections.)

B. Comparison Systems

In addition to the experimental OTE and ITE hearing aids described in Chapter II, five additional sound-reproduction systems were included in the Master Comparison Tape recordings. These are discussed below.

In order to obtain an estimate of the differences among good high fidelity loudspeakers, a pair of Electro Voice Sentry V loudspeakers was obtained. These are popular high-efficiency two-way studio monitor loudspeakers, which are acoustically equivalent to the EV Interface C consumer-product loudspeakers. The latter obtained an accuracy score of 87 percent in recent tests of high-priced "state-of-the-art" loudspeakers (Consumers Reports, 1978). Their principle inaccuracy is a dip in their power-response curve near the 2-kHz crossover frequency.

The Koss PRO4AA was selected as an example of a popular stereo headphone. This headphone was designed to produce a wide bandwidth with (intentionally) exaggerated bass response, a design which presumably accounts for it's popularity in hifi dealers' showrooms. (Koss produces more accurate headphones, but anecdotal evidence indicated that the PRO4AA was the largest-selling headphone in the U.S. until

superceded by the PRO4AAA model.)

A speech audiometer was simulated by using a pair of TDH-39 earphones (in MX-41/AR cushions) which were factory selected to have a frequency response nearly identical to the published "typical" response curve. A commercial speech audiometer was not used because of the ±5 dB frequency-response inaccuracies and the five- to ten-percent equivalent Total Harmonic Distortion at +6 dB VU (no other distortion test is specified) allowed by ANSI Standard S3.6-1969. Rather, the same professional quality amplifier and tape reproducer used with the reference system were used in order to provide an essentially flawless "speech audiometer."

As a representative from the low end of the range of systems advertised as "high fidelity", a Soundesign Model 6024 stereo phonograph (typically sold at discount department stores) was included in the comparisons. This "Discount Stereo" model consists of a record changer and amplifier in one module, with two separate (presumably single-element) loudspeakers.

Finally, a GE pocket radio (purchased in 1976 for $4.95) was included to serve as a low-fidelity anchor for the fidelity-rating scale. This pocket radio is a marvel of cost-effective engineering, but high fidelity it isn't.

The abbreviated designations for these seven comparison systems are as follows:

1. Pocket Radio (PR)
2. Discount Stereo (DS)
3. Speech Audiometer (SA)
4. Popular head Phones (PP)

5. Monitor Speakers (MS)

6. In-The-Ear hearing aids (ITE)

7. Over-The-Ear hearing aids (OTE)

C. Comparison Recordings: Four-track Master Comparison Tapes

1. Reference system response

The Master Stimulus Tape was reproduced on a two-track Ampex 440 professional tape recorder whose output was fed through a 125-watts-per-channel Marantz 250 stereo amplifier to a pair of Acoustic Research AR3a loudspeakers spaced along one wall of a 170 m^3 (6000 cu ft) room in the Auditory Research Laboratories of Northwestern University. The room dimensions were 6.1 m by 7.6 m by 3.7 m high. The sound absorption treatment on the walls and floor of that room was adjusted to eliminate audible flutter echos and to provide the .3 to .5 second reverberation time typically recommended for recording studios of that volume (see, for example, Olson, 1967). The resulting "room response" of the AR3a reference system is shown in Figure 4-1a. This response curve was obtained with one-third-octave bands of noise (uncorrelated) fed simultaneously to both loudspeakers. The resulting sound field was sensed with a small omnidirectional microphone having flat random-incidence response. The microphone was located symmetrically between the two loudspeakers and 3.3 m back from the 6.1 m wide wall along which the loudspeakers were placed. The optimum room response of a loudspeaker system is still a controversial subject, but the response shown in Figure 4-1a is quite similar to that recommended by Allison and Berkovitz (1972), Møller (1974) and Bevin (1978) based on
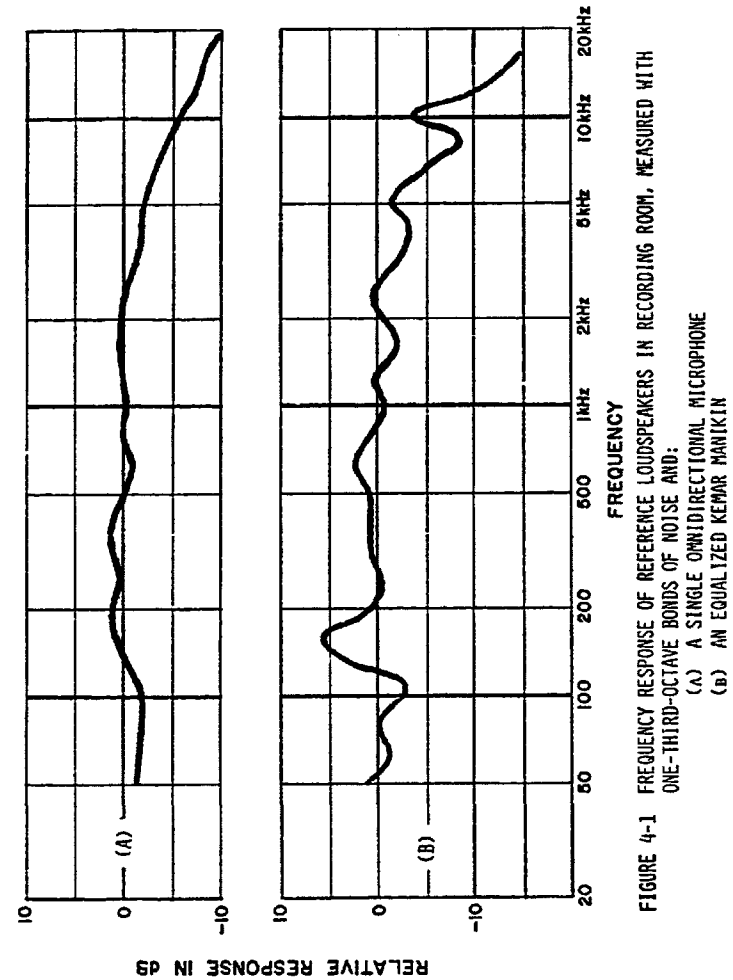
FIGURE 4-1 FREQUENCY RESPONSE OF REFERENCE LOUDSPEAKERS IN RECORDING ROOM, MEASURED WITH ONE-THIRD-OCTAVE BONDS OF NOISE AND:
(A) A SINGLE OMNIDIRECTIONAL MICROPHONE
(B) AN EQUALIZED KEMAR MANIKIN

listening-test results and a consideration of the high-frequency roll-off found in concert halls at typical listening locations compared to typical recording-microphone locations.

2. Manikin equalization

With the Master Stimulus Tape reproduction as source, binaural Master Comparison Tapes were recorded from the output of eardrum-position microphones in the KEMAR manikin after equalization to remove the spectral peak of approximately 15 dB at 2700 Hz produced by the external-ear resonances (a description of the bridged-T equalization filter and its rationale is given in Killion, 1979a). The manikin was placed one meter to the right of the room midline and 3.3 m from the wall along which the AR3a reference loudspeakers were located. The frequency response of the reference system as sensed by the equalized manikin microphones is shown in Figure 4-1b. This response curve was obtained with one-third-octave bands of noise as described above, except the response was taken as the average (on a power basis) of the manikin's left- and right-ear equalized microphone outputs. The increased unevenness seen in the manikin-sensed response was caused by imperfect equalization of the manikin at high frequencies and an unfortunate choice of manikin location at low frequencies.

3. Reproduce levels

As measured on the "C-weighting, fast" settings of a sound-level meter, peak levels during reproduction were 95 dB for the fortissimo Beethoven passage, 90 dB for the forte passage, 84 dB for both of the piano trio selections, 82 dB for the wideband noise, and 72-74 dB

for the male voice. These levels were chosen to duplicate as closely as possible the "live" levels a listener would have experienced at the original performance. (The level for the fortissimo orchestral passage was chosen equal to that commonly reported for live performances, as discussed earlier in Chapter II (page 60). The levels for the piano trio selections were chosen on the basis of the writer's own experience as an amateur jazz pianist. The level for the wideband noise was chosen arbitrarily.) Except for the loudspeaker-loudspeaker comparisons, a true live-voice source (the writer seated 1 m in front of the manikin and watching a sound level meter) was substituted for the pre-recorded voice selection on the Master Stimulus Tape.

4. Comparison conditions

The hearing-aid comparison recordings were obtained under exactly the same reproducing and recording conditions used for the reference recordings except that the OTE or ITE hearing aids were placed on the manikin and adjusted to unity insertion gain as described above (Chapter II) and below (this section).

The comparison loudspeaker recordings (MS and DS) were obtained with the loudspeakers substituted for (placed on the same one-meter-high stands previously occupied by) the AR3a reference loudspeakers. The Monitor Loudspeakers had "high-frequency rolloff" controls which were set to the position marked "flat".

The headphones were adjusted on the KEMAR manikin--with the help of tape and discs of closed-cell foam--to produce as close as could be estimated the equivalent of a real-ear seal and/or pinna deformation.

In the case of the TDH-39/MX-41AR headphones, the low-frequency attenuation due to the well-known leak around the ear cushions was made equal to the average obtained from probe-tube measurements on real ears as given by Shaw (1966) and confirmed by Killion, Tillman and Young (unpublished probe-tube data obtained on six subjects). Between 200 and 10,000 Hz, the resulting "eardrum-pressure" response measured on the unequalized KEMAR manikin fell within 2-4 dB of the predicted real-ear response calculated from Shaw's data for a typical TDH-39/MX-41AR earphone.

The pocket radio was located in the pocket of a shirt worn by the manikin. With the exception of the Discount Stereo (DS) and Pocket Radio (PR) systems, all loudspeakers and headphones were driven from the output of the same Marantz 250 stereo amplifier used with the reference loudspeakers. The headphones were driven through a 20 dB passive attenuator with 10 Ohm output impedance, an attenuator required to bring the 125 watt amplifier outputs down to suitable earphone-drive levels. The amplifiers in the Discount Stereo and Pocket Radio were included in the listening-test recordings of those two systems. Both amplifiers produced noticable distortion at high levels.

The gain of each sound-reproduction system was adjusted to produce the same levels on the speech-spectrum noise (82 dB SPL) and the Beethoven passage peak (95 dB SPL), as determined with a VU meter monitoring the equalized eardrum-position microphone outputs. When the frequency response of the comparison was substantially different from the reference system, it was sometimes impossible to simultaneously meet both level requirements and/or equal loudness (for the

writer's ears) between the reference and comparison system for all program selections. In those cases, a compromise adjustment was made. This compromise adjustment was required for the TDH-39 earphones (because of their minimal low-frequency response with the MX41-AR cushions) and for the Popular Phones (because of their exaggerated low-frequency response).

5. Comparison systems' responses

The relative frequency responses of six of the seven comparison systems are shown in Figure 4-2. Each response was obtained by subtracting the manikin-sensed reference curve of Figure 4-1b from the manikin-sensed response of the sound system under test. The latter was obtained with one-third-octave bands of noise in essentially the same manner used to obtain the reference curve. Thus, the hearing aid response curves (ITE and OTE) in Figure 4-2 are insertion-gain curves of those hearing aids on the KEMAR manikin. The remaining response curves represent difference curves, and reflect only the accuracy to which the system under test could duplicate the response of the reference high-fidelity system. By oversight, no frequency response was obtained for the Discount Stereo system.

D. Listening-Test Recordings: Binaural Listening Comparison Tapes

The ABA comparisons were organized into six program-selection blocks each containing seven system-comparison units. Each unit consisted of a spoken comparison-identification number, a five-second (approximately) segment from the reference system (A) recording, a five-second segment of the comparison system (B) recording, and

FIGURE 4-2   FREQUENCY RESPONSES OF SOUND SYSTEMS USED IN LISTENING TESTS.

NOTE: HEARING AID RESPONSES ARE INSERTION GAINS; ALL OTHERS ARE RELA-
TIVE TO REFERENCE LOUDSPEAKER SYSTEM RESPONSE.

131

another five-second segment of the reference (A). The same ABA com-
parison was repeated in order to permit a "second listen" to each
comparison. Including pauses, each complete unit occupied about
40 seconds. (The total of 6 x 7 = 42 comparison units occupied just
under 30 minutes after the program-selection-block announcements were
included.)

The ABA comparisons were recorded on a two-track Ampex 440 re-
corder from either (A) tracks 1 and 2 or (B) tracks 3 and 4 of the
four-track Master Comparison Tape mounted on an Ampex 440 reproducer.
Since the track pairs had been synchronized during the original
Master Comparison Tape recording sessions, the ABA comparisons were
obtained by simply switching between track pairs at appropriate times.
In the case of the four musical passages, therefore, a continuous
musical passage--the middle portion of which had been reproduced over
the comparison system--was recorded. (Within each musical-selection
block, an attempt was made to hold the switchover points to the same
beat of the same measure for all comparisons.) In the case of the
live voice, tne same "Joe ... lawn." nonsense sentence was recorded
three times, the second time from the comparison-system reproduction
of the previous anechoic-chamber live-voice recording.

Within each of the six program-selection blocks, the first
comparison unit was always for the low-fidelity Pocket Radio, but the
remaining six comparisons were randomized according to a Latin-Square
form of randomized-block design. (Thus, each of the six nominally-
high-fidelity sound systems was represented once in every position in
the presentation order.)

E. A Flaw

One speaker problem occurred: The supertweeter on the right
AR3a reference speaker opened up before the experiment was underway
(perhaps due to ultrasonic oscillations in the power amplifier before
the system wiring was cleaned up). This tweeter was "repaired" by
resoldering a burned-out lead wire. The repair survived the initial
frequency response checks, but near the end of the recording session
it was discovered that the repair had failed near the beginning of
the comparison recordings.

The majority of the audible energy in the program selections above
8 kHz (where the supertweeter came into play) was contained in the
left (drummer) channel of the piano trio selections, which explains
how the tweeter failure escaped detection through most of the record-
ing sessions.

After auditing several new-vs-old comparisons, the writer con-
cluded that the small change in sound quality did not warrant re-
doing several weeks of recorded comparisons. (On an energy basis,
the failure would have caused approximately a 3 dB reduction in level
above 8 kHz if there had been equal high-frequency energy in each
channel. Since the left channel contained most of the high-frequency
energy and the right supertweeter failed, the level change was pre-
sumably much less than 3 dB.)

Nonetheless, that tweeter failure marred what the writer felt
was an otherwise extremely careful fidelity-rating experiment. For
this reason, an "extra comparison" (so identified on the tapes) was
included, using the Monitor Speakers and the repaired AR3a's on the

fortissimo Philharmonic passage, to provide a gross indication of the
importance of the tweeter failure.

F. Subjects and Procedures

In order to obtain as close as possible to a "man-on-the street"
jury, a group of 24 Untrained Listeners was selected by the personnel
department of a manufacturing concern to meet only the following cri-
teria: equal male-female representation, approximately rectangular
age distribution between age 20 and age 60, and as wide a distribution
of occupations as could be obtained. (The final criterion was included
to avoid the possibility of obtaining a heavy technical representation
on the listening jury). The resulting jury contained 12 males and
12 females, with eight subjects in their twenties, five in their
thirties, four in their forties, six in their fifties and one 61 year
old subject. No hearing tests were performed on any subjects for the
purpose of this experiment.

The Untrained Listeners were made available for two one-hour
sessions on successive days. On each day, the subjects rated nine
blocks of seven comparisons. The first three blocks were from
"Comparison Tape A" (described below) on the first day and "Compari-
son Tape B" (described below) on the second day, followed by the six
program-selection blocks of the main experiment (described above).
The 21 comparisons of the three initial blocks were treated as prac-
tice comparisons for the purposes of the main experiment, although
the subjects were not so informed. The order of the six blocks of
the main experiment was randomized (by rewinding tape) for the second

To give you an example, if the program material was shaped wide-band noise (as from a distant waterfall), the reference system might sound like "shhhh." If the comparison system produced a sound like "sssss", you would hear shhhh - sssss - shhhh, and then again shhhh - sssss - shhhh. In that example case, I would expect you to give a relatively low fidelity rating to the system under test, because "shhhh" and "sssss" sound quite a bit different. On the other hand, if you heard shhhh - shhhh - shhhh, shhhh - shhhh - shhhh, I would expect you to assign a fidelity rating near 100 percent because the sound was very similar. Note that you are not asked to make sound quality judgments, but similarity judgments. Thus if the sound in the middle section sounds different, you should give it a rating less than 100 percent even if you think it sounds better. There may be differences in loudness, but try to ignore them.

The comparisons come in blocks of seven. The first comparison you will hear in each block is for a low-quality system. I expect you will give it the lowest rating of any of the systems.

II. RESULTS AND DISCUSSION: THE MAIN
24-SUBJECT EXPERIMENT

A. Average Fidelity Ratings

The mean fidelity ratings averaged over all subjects and the six program selections are shown in Figure 4-3. Both the OTE and the ITE hearing aids obtained higher ratings than any of the other high-fidelity systems. The standard error of the mean ratings—based on
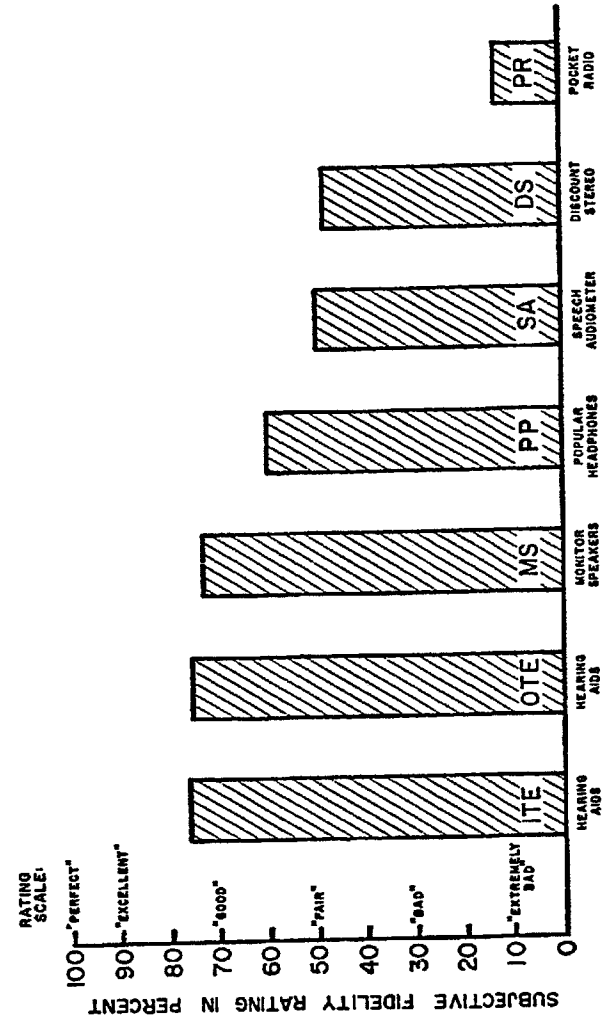
FIGURE 4-3 AVERAGE FIDELITY (SIMILARITY) RATINGS FOR SIX PROGRAM SELECTIONS FROM MAIN EXPERIMENT USING 24 UNTRAINED LISTENERS.

the system-subject interaction obtained from a three-way analysis of variance--was less than 1.6 percent. A t-test applied to the differences between the hearing aids and the other systems indicated no significant difference between either of the hearing aids and the Monitor Speakers. All other differences between the hearing aids and the other systems were significant at the .000001 level or better (the smallest of those differences was ten times the standard error of the mean). Thus, null Hypotheses H_5 and H_6 were rejected.

Stated somewhat differently, the change in sound quality caused by interposing either pair of hearing aids in the sound path between the reference loudspeakers and the eardrum-position microphones in the KEMAR manikin was rated comparable to the change in sound quality caused by changing from the AR3a reference loudspeakers to a different pair of high-quality loudspeakers. The change in sound quality caused by interposing either pair of hearing aids was judged to be significantly _less_ than that caused by changing from the reference loudspeakers to (the amplifier and speakers from) a Discount Stereo phonograph, Popular Phones, a Speech Audiometer, or (not surprisingly) a Pocket Radio.

These results may appear surprising to those familiar with the design compromises found in conventional hearing aids, although they are entirely consistent with the objective data presented in Chapter II. Recall, for example, that the calculated 21-band accuracy score for both the OTE and ITE hearing aids fell in the upper half of the

range of scores obtained by inexpensive and high-priced (respectively) high-fidelity loudspeakers tested recently at Consumers Union. Given that result, it is not surprising that the hearing aids rated significantly higher than the Popular Phones with their exaggerated bass response, or the Speech Audiometer with the severe bass loss produced by the well-known cushion leak. Both defects were readily apparent in the frequency response curves shown in Figure 4-2. Although an objective measure of the frequency response of the "Discount Stereo" system was not obtained, subject comments (optional) indicated it had a "high-frequency rolloff", a "mid-frequency dip", a "hollow sound", and a "lack of bass response." (In the writer's judgment, it also had a "boomy" mid bass and it distorted badly on the fortissimo orchestral passage.)

The fact that the OTE aids with only 8 kHz bandwidth rated as well as the ITE aids with 16 kHz bandwidth was presumably due to the comparable importance of the different defects in their frequency response. The OTE aids had a limited bandwidth but an extremely smooth insertion-gain frequency response, while the ITE aids had a sensibly unlimited bandwidth but a dip in response near 2.7 kHz due to their imperfect compensation for loss of normal external-ear resonances. The high rating of the OTE aids came as somewhat of a surprise to the writer, although it was entirely consistent with the conclusion reached by Fletcher (1942) that "substantially complete fidelity (for) ...orchestral music is obtained (with)...a frequency range of from

60 to 8000 cycles per second." The average rating for the OTE aids on the two orchestral passages was 85 percent (see Appendix B). Snow (1931) reported a value of 91 percent obtained in A-B-A-B-- quality-rating comparisons using orchestral music for a system with an 8 kHz upper cutoff frequency and no other defects. To the extent the two experiments are comparable, the out-of-ear microphone location with the OTE aids appears not to have been a major defect.

Interestingly enough, the 73 percent average rating for the high-quality monitor loudspeakers (and the 75 and 76 percent ratings for the experimental hearing aids) was almost exactly equal to the 74 and 75 equivalent percentage of the 7.4 and 7.5 decimal ratings obtained for two "high" fidelity loudspeakers by Gabrielsson, Rosenberg, and Sjögren (1974) using a similar rating scale but the subject's memory of how a live performance sounded as the reference.

Despite the apparent reasonableness of the results from the main experiment, it is within the realm of possibility that the ratings shown in Figure 4-3 were peculiar to the selection of program materials or subjects, or to the reproduce conditions under which the comparisons were presented. Several additional experiments were undertaken to examine those possibilities. The remainder of this chapter is devoted to an examination of the data from the main experiment and the additional experiments, in an attempt to estimate the reliability and validity of the results described above.

B. Test-Retest Reliability

Since all comparisons in the main experiment were repeated (in different program-selection-block order) on the second day of testing, a comparison of the two day's ratings provides not only a statistical reliability indication, but an indication of the importance of learning, seating position (recall that the two loudspeaker channels were reversed for the second day's comparisons), and signal-to-noise ratio during comparison presentation.

Figure 4-4 shows a graphical comparison between the two day's ratings, based on the data presented thereafter in Table II. The calculated correlation (Pearson's product-moment) coefficient between the two sets of mean ratings was extremely high; $r = .998$. (All correlation coefficients discussed in this chapter were found to be significant at well beyond the .01 level, based on an F-test applied to the linear regression analysis.) These results indicate that the ratings were relatively independent of the factors listed above.

Indeed, additional data were obtained from nine of the writer's relatives who were imposed upon to "take the listening test" during visits to his home. These comparisons were reproduced over an old (relatively low-fidelity) "hi-fi" in the writer's home, at levels which were estimated to range between 5 and 15 dB below those used in the main experiment. The correlation (.997) between those ratings and the average ratings from the main experiment was as good as the test-retest correlation obtained in the main experiment.

As previously discussed in Chapter III, these results have a certain face validity: The fidelity ratings obtained in the present
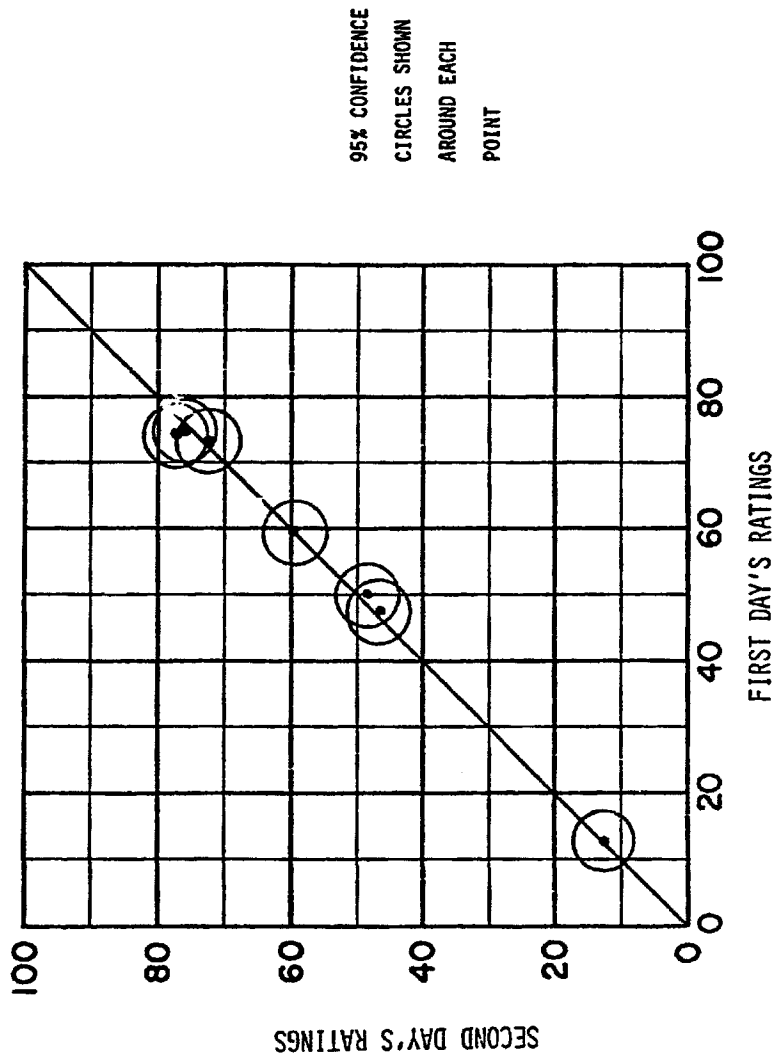
FIGURE 4-4 COMPARISON BETWEEN FIDELITY RATINGS OBTAINED ON TWO DIFFERENT DAYS

FIRST DAY'S RATINGS

SECOND DAY'S RATINGS

TABLE II

OVERALL FIDELITY RATINGS FOR SEVEN SOUND SYSTEMS
OBTAINED FROM 24 UNTRAINED LISTENERS

Sound System	First Day	Fidelity Ratings Second Day	Average
ITE Hearing Aids	74.2	77.3	75.7
OTE Hearing Aids	74.7	76.0	75.4
Monitor Speakers	73.1	72.3	72.7
Popular Head Phones	59.2	59.5	59.3
Speech Audiometer	50.0	48.4	49.2
Discount Stereo	47.3	46.8	47.1
Pocket Radio	12.5	12.8	12.6

experiments were basically _similarity_ ratings, as stated in the instructions to the subjects. Thus, the _constant_ abberations in sound quality introduced by any reasonable sound reproduction system might be expected to have little effect on a subject's ability to detect _changes_ in sound quality between two segments of a pre-recorded comparison.

C. Three-way Analysis of Variance

A three-way analysis of variance was performed on the 2016 individual-subject ratings obtained in the main experiment. The complete analysis is presented in Appendix B. Application of the F-test to the results (see Table III) indicated there were indeed significant differences among sound systems, program materials and subjects; and statistically significant interactions between each. Only the three-way system-program-subject interaction was not significant. (Indeed, all other differences were statistically significant at well beyond the .001 level.) Thus, null Hypotheses H_1 through H_4 were rejected.

At the same time, finding a _statistically_ significant interaction with the large data base available here does not imply, per se, an interaction which has any _practical_ significance. On an a priori basis, only the system-program interaction might be expected to have practical importance. A discussion of the effect of program selection on the individual system ratings will be deferred to a later section, however, so that additional subject-group data may be included.

The effect of applying the arcsin transformation to the raw data before performing the analysis of variance was investigated (see Appendix B). As expected, that transformation did stabilize the

TABLE III

SUMMARY OF ANALYSIS OF VARIANCE PERFORMED ON MAIN FIDELITY-RATING EXPERIMENT WITH SEVEN SYSTEMS, SIX PROGRAMS, TWENTY-FOUR SUBJECTS, AND TWO REPLICATIONS

Source of Variation		Sum of Squares	Degrees of Freedom	Mean Squares	Error Term	F	Critical F .01	Critical F .001
Sound Systems	A	881629	6	146938	AC	203.5	2.96	3.7
Program Material	B	127825	5	25565	BC	31.1	3.17	4.4
Subjects	C	125368	23	5451	ERR	22.4	2.0	2.3
System x Program	AB	95620	30	3187	ERR	13.1	1.8	2.2
System x Subject	AC	99674	138	722	ERR	3.0	1.4	1.7
Program x Subject	BC	94578	115	822	ERR	3.4	1.4	1.7
System x Program x Subject	ABC	201601	690	292	ERR	1.2	1.3	1.5
Replication	ERR.	245285	1008	243				

variance of the Untrained-Listener ratings, but it had negligible effect on either the average ratings or the significance level of any comparisons. (This was largely due to the fact that none of the average ratings was near 100% or 0%). All results reported here were obtained from the intuitively simpler untransformed data.

III. ADDITIONAL EXPERIMENTS AND ANALYSES

A. Trained-Subject Ratings

1. Subjects and procedures

A second group of "Golden Ears" subjects was enlisted. This group consisted of five individuals (Alf Gabrielsson, Julian Hirsch, Hugh Knowles, Bruno Staffen, and Edgar Villchur) each of whom had devoted a large amount of time at some point in his life to the subjective evaluation of high-fidelity loudspeaker systems.

A third group of Chicago-area "Trained Listeners" was enlisted. This group consisted of six individuals (Elmer Carlson, Richard Peters, Daniel Queen, Eugene Ring, Robert Schulein, and Frederic Wightman) each of whom had considerable training in listening experiments, although not necessarily in high-fidelity-system evaluations.

The Golden-Ear and Trained-Listener groups were sent copies of the instructions and listening-test tapes (the tape copies were made by Webb Recording, Chicago) and asked to use their best headphones during their evaluations. They were further instructed to set the headphone levels for the equivalent of an 84 dB sound field while reproducing a calibration segment of 84 dB SPL speech spectrum noise.

Each listener was given different instructions as to the order in which he was to listen to the tapes, but no further attempt was made to counterbalance presentation order.

2. Average ratings

The average ratings obtained from the five Golden-Ear subjects and the six Trained-Listener subjects are given in Table IV. They are also shown (inset for comparison) in Figure 4-5, which is otherwise a duplicate of Figure 4-3. The average ratings are qualitatively quite similar to those obtained from the Untrained-Listener subjects, an observation which will be discussed quantitatively after consideration of the three-way analysis of variance performed on each group's data.

3. Analysis of variance for trained-subject data

An analysis of Variance (see Appendix C) applied to the Golden-Ear and the Trained-Listener data produced the same conclusions as stated above for the Untrained-Listener data, with the exceptions and observations noted below:

a) The error variance (estimated from the three-way interaction between systems, programs, and subjects) for both the Golden-Ear and Trained-Listener subjects was nearly four times smaller than that for the Untrained-Listener subjects. Not surprisingly, highly trained listeners are much more consistent in making subjective judgments than untrained listeners.

b) The system-subject interaction was not significant for the Trained-Listener subjects, indicating a high degree of

TABLE IV

OVERALL FIDELITY RATINGS FOR SEVEN SOUND SYSTEMS OBTAINED FROM THREE SUBJECT GROUPS

Sound System	Subject Group		
	Untrained Listeners (N=24)	Golden Ears (N=5)	Trained Listeners (N=6)
ITE Hearing Aids	75.7	74.6	86.2
OTE Hearing Aids	75.4	75.9	77.9
Monitor Speakers	72.7	67.5	79.9
Popular Phones	59.3	43.5	57.5
Speech Audiometer	49.2	42.2	54.5
Discount Stereo	47.1	42.5	56.6
Pocket Radio	12.6	17.7	14.5
Average	56.0	52.0	61.0

149
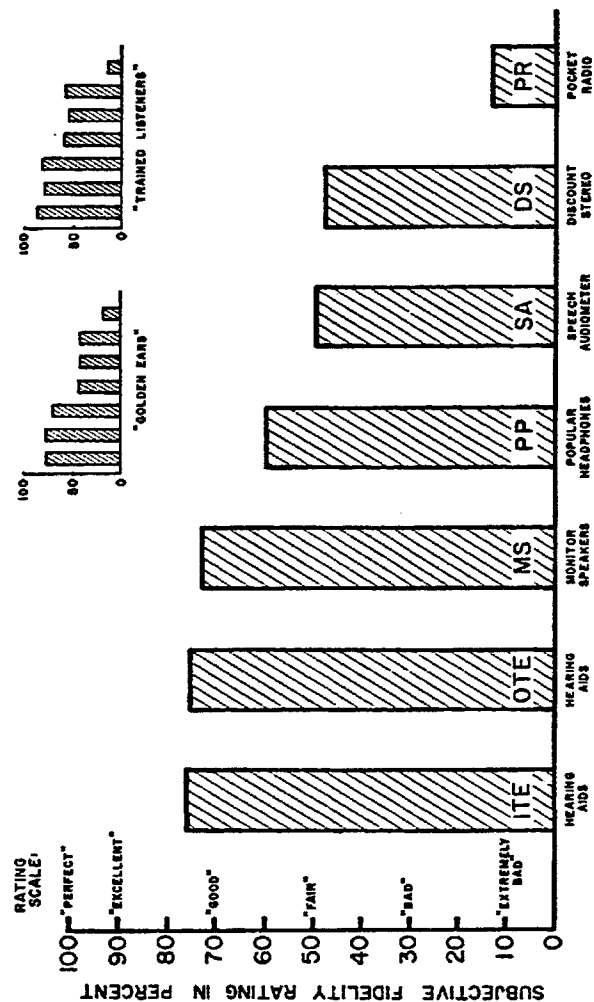
FIGURE 4-5   AVERAGE FIDELITY (SIMILARITY) RATINGS FOR SIX PROGRAM SELECTIONS FROM MAIN EXPERIMENT USING 24 UNTRAINED LISTENERS. RESULTS FROM 5 "GOLDEN EARS" AND 6 "TRAINED LISTENERS" SHOWN INSET.

150

homogeneity among that group. (All were known to have spent
an appreciable amount of time listening to and/or performing
music, an observation which may or may not be relevant).

c) The standard error of the mean, estimated from the variance
due to system-subject interaction, was 2.5 percent for the
five Golden-Ears subjects and 1.6 percent for the six Trained-
Listener subjects. The greatly reduced variance exhibited
by the two trained-subject groups meant that the reliability
of their single-session average ratings was comparable to
that obtained from two sessions with the much larger (N = 24)
group of Untrained Listeners. In particular, one trained
subject appears to be worth at least eight untrained sub-
jects where statistical sample-size considerations are con-
cerned. This hardly surprising result is qualitatively
similar to that obtained by Gabrielsson and Sjogren (1976).

d) The variance due to system-program interaction was almost
20 times smaller for the Golden-Ear subjects, and nearly
10 times smaller for the Trained-Listener subjects, than
for the Untrained-Listener subjects. Successful Golden-Ear
professionals have presumably found it useful to train
themselves to "listen through" the particular music selec-
tion used for system evaluation. Although the program
selections were considered as "fixed effects" in the sta-
tistical analysis of the present experiments, the calculated
reliability of the trained-subject ratings would have suf-
fered relatively little if the program selections had been

considered a random sample. In other words, essentially
similar ratings might be expected from trained subjects
using any reasonable cross-section of program material.

4. Comparison to Untrained-Listener results

No generally accepted solution to the comparison of the means of
two unequal samples from two populations whose variances may be un-
equal (as they clearly are in this case) exists, but Welch's approxi-
mation is often used (Brownlee, 1965; p. 299). The application of
this t-test method to the differences between the overall average
ratings obtained from the Untrained-Listener and the Golden-Ear and
Trained-Listener subjects indicated the differences were not signifi-
cant at the .05 level.

Application of the Bonferroni inequality and t-statistics
(Miller, 1966) to obtain confidence intervals for the seven individ-
ual system ratings from each group, however, indicated the 95 percent
confidence intervals were ± 4.6 percent, ± 7.8 percent, and ± 4.8
percent for the Untrained-Listener, Golden-Ear, and Trained-Listener
groups, respectively. Thus some of the between-group differences
in individual system ratings appeared to be significant. The most
striking was the roughly 15 percent lower rating given the Popular
Phones by the Golden-Ear subjects compared to the two other subject
groups. That seemed reasonable in light of the comment made by one
of the Untrained-Listener subjects, who ignored instructions and gave
the Popular Phones a 100 percent rating because he "liked them much
better" (than the reference). In particular, anecdotal market

evidence indicates that those who have not spent a lot of time profes-
sionally evaluating high-fidelity systems are much more tolerant of an
excessive bass response than a deficient bass response.

The correlations between the Golden-Ear and Untrained-Listener
ratings (r = .956) and the Trained-Listener and the Untrained-
Listener ratings (r = .984) were both high. This provided further
evidence of the stability of the relative ratings produced in the
present experimental design, since the Untrained Listeners heard the
comparisons reproduced over loudspeakers, whereas the Golden-Ear and
Trained-Listener subjects heard the comparisons over (their own
favorite) headphones. The good relative agreement between Trained-
and Untrained-Listener ratings is consistent with the findings re-
ported by Gabrielsson, Rosenberg and Sjögren (1974) and Gabrielsson
and Sjögren (1976).

NOTE

In comparing the high correlation coefficient (the Pearson pro-
duct-moment correlation coefficient r has been used throughout) to
the obvious differences among ratings from the different subject
groups, it should be recalled that the correlation indicates the
degree to which the least-squares best-fit linear relationship (y =
mx + b) accounts for the dependent-variable data. In particular,
r^2 can be regarded as measuring the fraction of the variance in the
dependent-variable data "explained" by the regression on the
independent-variable data, and vice versa (Brownlee, 1965). After
accounting for the differences between Untrained-Listener and

Trained-Listener ratings (by applying the "optimum" linear transforma-
tion from one to the other), for example, all but $1-(.984)^2 = .03$
(three percent) of the variance is accounted for. Simply stated, both
groups appear to be measuring essentially the same thing using slightly
different subjective scales.

B. Effect of Program Selection

Although the program materials were treated as fixed effects in
the analysis of variance, it is reasonable to ask how the ratings
would have turned out if a different selection of material had been
used in the comparisons. One indication of this answer can be
obtained from an examination of the ratings on the individual selec-
tions. These ratings are shown plotted in Figure 4-6 and are given
in Table V, where the three mean ratings (one from each subject
group) are averaged for each of the four types of program materials:
orchestra (two selections); piano trio (two selections); live voice
(one selection); and wideband noise (one selection). Also shown are
the grand-mean ratings obtained from this averaging procedure. Note
that each subject group and each type of program selection is given
equal weight in these grand-mean ratings. Simple averaging across
subject groups appeared justified since the eight linear correlation
coefficients obtained by comparing the Golden-Ear and Trained-
Listener ratings for each type of program selection to the Untrained-
Listener ratings were all high (average r = .96; range from .90 to .99).

The type of program selection clearly affects the absolute values
of the ratings. (Recall that hypothesis H_2--that there were no
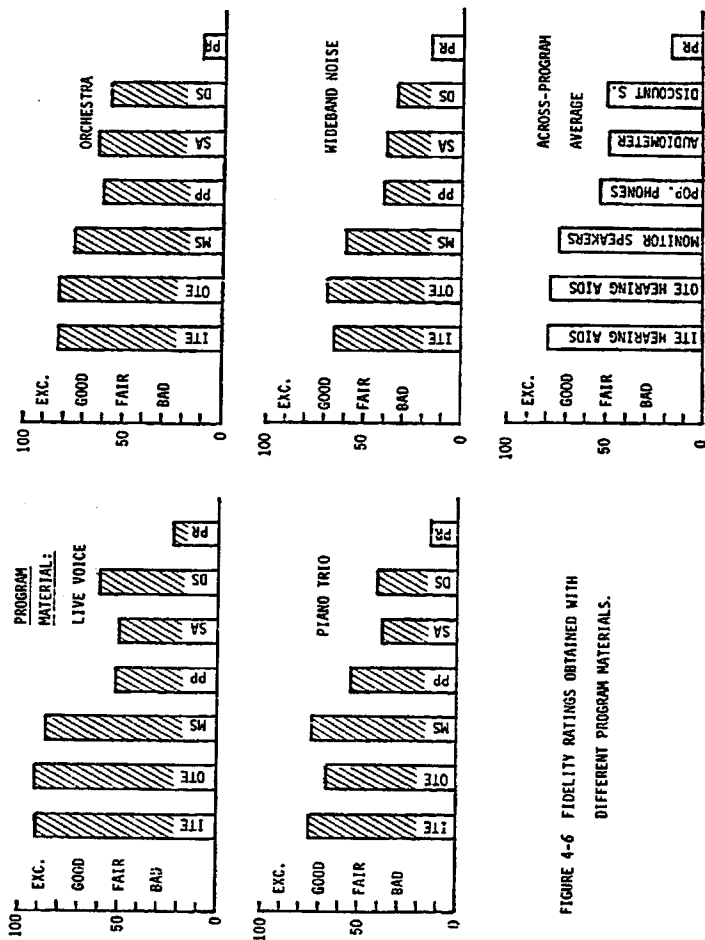
FIGURE 4-6 FIDELITY RATINGS OBTAINED WITH DIFFERENT PROGRAM MATERIALS.

TABLE V

ACROSS-SUBJECT-GROUP FIDELITY RATINGS FOR FOUR
TYPES OF PROGRAM MATERIALS

Sound System	Program Material				Across-Program Average
	Live Voice	Orchestra	Piano Trio	Wideband Noise	
ITE Hearing Aids	91.2	82.8	75.3	65.7	78.8
OTE Hearing Aids	91.9	82.4	66.7	68.4	77.4
Monitor Speakers	86.7	75.0	74.3	54.8	72.7
Popular Phones	50.8	61.1	53.8	40.2	51.6
Speech Audiometer	49.3	63.2	38.6	39.4	47.6
Discount Stereo	59.9	57.8	41.7	34.2	48.4
Pocket Radio	22.7	11.2	14.2	16.2	16.1

Note: Ratings for orchestra and piano trio are averages for two
different musical passages. Thus, the across-program average
ratings shown here differ slightly from the six-program
averages shown in Tables III and IV.

differences among ratings produced with different program selections--
was easily rejected on the basis of the analysis of variance of the
Untrained-Listener data. The same conclusion results from the same
analysis applied to the other two subject-group data sets.) On the
basis of the grand-means across sound systems and subject groups, the
Live Voice is easiest to reproduce (64.6%), followed by the symphony
orchestra (61.9%), the piano trio (52.1%), and wideband noise (45.6%).

The most obvious interaction evident in Figure 4-6 was the ex-
pected one: the OTE hearing aids were rated relatively lower on the
piano trio selections where the audible very-high-frequency energy
produced by the drummer's cymbols presumably made the 8 kHz cutoff
frequency of the OTE hearing aid obvious. The relatively poor rating
for the speech audiometer on the piano trio selections was expected
because the poor low-frequency response caused by the leak around the
MX-41AR earcushion produces a very thin bass sound. The poor rating
for the speech audiometer on the live voice comparisons was not
expected. The explanation is partly the same (the male voice has
audible energy down to 100 Hz or so, well below the roughly 500 Hz
cutoff frequency of the speech audiometer), but the lack of (manikin)
listening-room reverberation was also particularly noticeable in the
live-voice versus earphone comparisons.

Despite the obvious interaction between system and program selec-
tion, in some cases, it is clear that the same basic conclusions would
have been obtained with each type of program material, i.e. the OTE
and ITE hearing clearly fall in the "high fidelity" category by any
program-selection measure.

C. Additional Sound-System Ratings

Comparisons of several additional sound systems were included in
the previously mentioned "practice tapes" A and B. These were includ-
ed for several reasons. One was to provide additional "benchmarks"
to better define the fidelity rating scale in terms of common experi-
ence. Another was to provide preliminary hearing-aid-design data on
the importance to fidelity judgments of undamped earmold resonances
and an In-The-Concha microphone location. A third was to provide ad-
ditional data to allow a better comparison (see section D, below)
between the subjective fidelity ratings and the calculated 21-band
accuracy score (which was based on objective frequency-response mea-
surements). A fourth was to provide a direct estimate of the fidelity
of the reference loudspeakers using a true "Live-vs-Recorded" condition
with live-voice source. And finally, some comparisons were included
simply to satisfy the writer's curiosity. Most of these additional
systems are described below.

1. Systems

Two additional hearing aid systems were included. One was the ITC
(In-The-Concha microphone) hearing aids with 16 kHz bandwidth which
were described in Chapter II. An additional system was obtained from
the OTE aids by simply changing from the "8CR" earmolds to conventional
earmolds consisting of 40 mm of #13 (1.9 mm) tubing extending from the
earhook to the tip of the earmold. The latter undamped-earmold condi-
tion was labeled "OTE-40".

The relocation of loudspeakers in a listening room is a common experience. A "Relocated Reference" (RR) system was obtained by simply moving the AR3a reference loudspeakers from their stands (located approximately 1 m off the floor and .75 m away from each side wall in the recording room) down into the corners of the recording room. At low frequencies, this amounts to changing from radiation into 2π steradians to radiation into $.5\pi$ steradians, a change which provides a 6-dB bass boost. In addition, the virtual images produced by the corner reflections can produce interference effects at midband frequencies. These two effects were discussed by Knowles (1941). As will be seen (Figure 4-7, p. 163), both acoustic effects were produced in the Relocated Reference condition. In addition to the Relocated Reference loudspeaker comparison, a pair of Scott S-10B two-way "Bookshelf Speakers" (BS) was included. Although these had substantially lower power-handling (and, more importantly, power-output) capability than either the AR3a or Sentry V speakers, they were able to (just) provide the 95 dB SPL orchestral peaks in the 170 m^3 recording room without audible (to the writer's ears) distortion. Their principal defect was a "beaming" at high frequencies with correspondingly reduced total output even with the high-frequency level adjustment on these speakers set to maximum (as it was for these comparisons).

Two additional headphone systems were included. One was the highly-rated Sennheiser HD-414X "open air" supra-aural headphones (HRP). The other was "Airline Phones" (AP) obtained from an EECO 169936-01 transducer, of the type normally mounted in the armrest of commercial passenger aircraft, fed through a common stethoscope-style

stereo headset. The HRP system was included because it produced the closest to a flat field-referenced frequency response of any (of the limited number of) headphones the writer had measured on the KEMAR manikin: The eardrum-pressure response measured on the KEMAR manikin with those earphones was quite similar to the eardrum-pressure response measured on the manikin in a diffuse sound field (Killion and Monser, 1979). The AP system was included because it appeared to be the closest thing to a universally-experienced "high-fidelity" system one might find. (No attempt was made to determine how representative the particular transducer chosen for these comparisons was of the several manufacturers' models in common use, however.)

Recall that all subjects were asked to attempt to ignore loudness differences in their fidelity (similarity) ratings, a request motivated by the writer's difficulty in unambiguously setting the appropriate comparison levels with systems whose frequency response was uneven or whose bandwidth was severely limited. Some measure of their success at this task seemed advisable. Thus, an additional "system" which comprised nothing more than a 3-dB increase in the level of the reference-system recording was included. This system was labeled "+3 dB".

2. Procedure

The same recording procedure used in the main experiment was used to obtain comparison recordings for the additional systems. Comparison Tape A contained three (randomized) program-selection blocks of seven system-comparison units each. The program selections were three of the six used in the main experiment; namely, the fortissimo

orchestral passage, the first piano-trio chorus, and the male voice. The seven sound systems were the Pocket Radio (the first comparison in each block), the OTE-40 hearing aids with undamped tubing, the Relocated Reference speakers (RR), the Bookshelf Speakers (BS), the Highly-Rated Phones (HRP), the Airline stereo Phones (AP), and the "+3 dB" level shift.

Tape B contained an assortment of comparisons in two program-selection (orchestra and piano trio) blocks. These included the ITC hearing aids, a Monaural condition, a switch to Larger Ears (Maxwell and Burkhard, 1978) on the KEMAR manikin, a Change in Microphones in the OTE hearing aids (from microphones with an 8 kHz cutoff to microphones with a 16 kHz cutoff, both with flat response below cutoff) to assess the possible effect on "transient response", and a "2:1 compression" condition obtained by simply inserting a high-quality compression amplifier in the OTE hearing aid amplifier chain. A final block on Tape B contained a series of true Live-Versus-Recorded comparisons, which are discussed in Section E, below.

The comparisons in the main experiment were recorded on two tapes labeled "C" and "D", so labeled because they were actually recorded after tapes A and B in order to provide the writer as much practice as possible before the comparisons for the main experiment were recorded. (Each of the four tapes lasted a little under 15 minutes, so they each occupied a $7\frac{1}{2}$ inch reel at the 15 ips recording speed used for all master tapes and for all but one of the copies sent to the trained subjects).

During the presentation of comparisons, the same instructions were used throughout. None of the subjects knew which tapes contained the "main experiment" and which contained the additional system comparisons, although tape A or B (depending on session and/or subject) always preceded tapes C and D in order to provide some practice before the main comparisons were rated. Except for practice effects, therefore, the ratings obtained from tapes A and B were directly comparable to the corresponding program-selection-block ratings from the main experiment. Appendix D contains a complete listing of all the comparisons on Tapes A, B, C and D, listed in the order they occur on each tape. (The average fidelity rating obtained from each subject group is also given for each comparison).

3. Results and discussion

The relative frequency response of most of the additional systems is shown in Figure 4-7, along with the previously-shown responses from the main experiment. As before, each response was obtained by subtracting the manikin-sensed reference curve of Figure 4-1b from the manikin-sensed response of the sound system under test. It is well to keep in mind that only the hearing-aid curves are "pure" (insertion-gain) frequency response curves. The remaining response curves represent difference curves, and reflect only the accuracy to which the system under test could duplicate the response of the reference high-fidelity system. Thus the rising response shown for the MS and HRP systems at high frequencies is the result of a relatively flat manikin-sensed field response for those systems at high frequencies from which
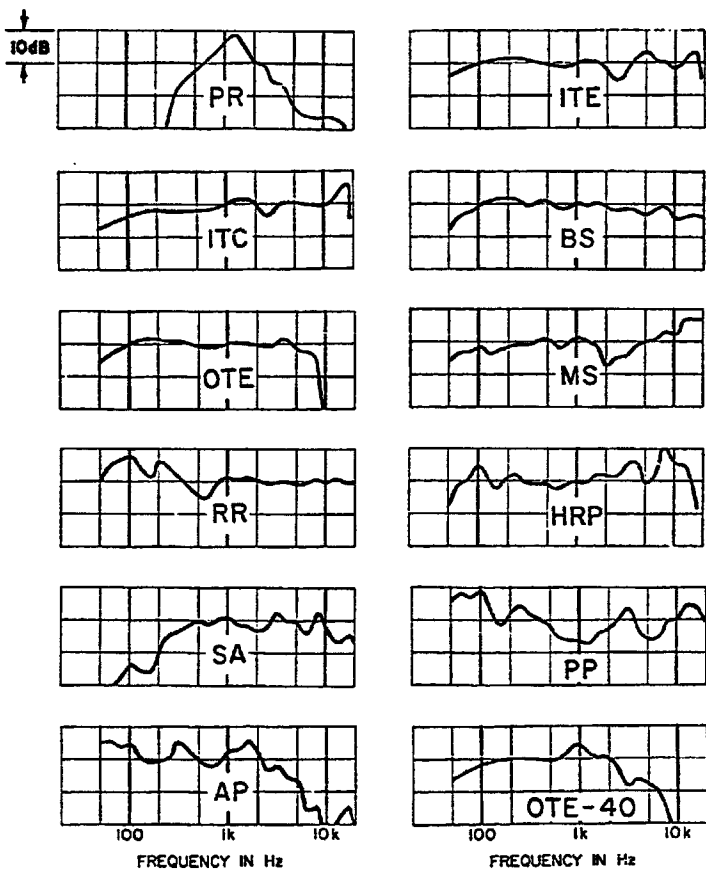
FIGURE 4-7 FREQUENCY RESPONSES OF SOUND SYSTEMS USED IN LISTENING TESTS.

NOTE: HEARING AID RESPONSES ARE INSERTION GAINS; ALL OTHERS ARE RELATIVE TO REFERENCE LOUDSPEAKER SYSTEM RESPONSE.

the (presumably desirable) moderate high-frequency rolloff seen in the AR3a Reference loudspeakers was subtracted. Allison and Berkovitz (1972) reported that the total-radiated-power response of the AR3a is relatively flat out to nearly 20 kHz, so that most of the high-frequency rolloff seen in the typical room-response of the AR3a is due to the greater high-frequency absorption in most rooms. This phenomenon may help explain the common observation that flat-response earphones such as the HD-414X sound too "bright" in listening tests (Toole, 1978). This problem is discussed further in Section E, below.

Figure 4-8 shows the average fidelity ratings given each of 17 sound-reproduction systems by the three subject groups. Table VI provides the same information in numerica form. These ratings were averaged over the three program selections on which all systems were compared. The standard error of the mean ratings ranged from 1.9 percent to 3.9 percent for these means, depending on the subject group and comparison blocks from which they were obtained. By simple computation, 95 percent of these ratings should be reliable within less than ±8 percent.

These ratings place the fidelity-rating results in better perspective: simply moving the same loudspeakers to a different place in a room can produce a sufficient change in sound quality to drop the ratings into the 85 percent region. This points out the importance of the well-known problem of A-B comparisons of speakers: even identical speakers won't sound the same unless they occupy exactly the same position in the room. Changing pinnae (a task more easily

FIGURE 4-8 AVERAGE FIDELITY RATINGS (ORCHESTRA, PIANO TRIO, LIVE VOICE) FOR 17 SOUND SYSTEMS

TABLE VI

FIDELITY RATINGS FOR SEVENTEEN SYSTEMS RATED ON THE SAME THREE PROGRAM SELECTIONS (ORCHESTRA, PIANO TRIO, AND LIVE VOICE)

Sound System	Subject Group			Listening Comparison Tape
	Untrained Listeners	Golden Ears	Trained Listeners	
Changed Mics.	87.5	92.4	96.4	B*
+3 dB Level Shift	83.7	88.0	95.9	A
Relocated Ref. Speakers	83.9	78.3	87.3	A
ITC Hearing Aids	79.6	88.3	90.0	B*
Monitor Speakers	79.6	73.3	85.8	CD
ITE Hearing Aids	78.1	77.5	89.9	CD
OTE Hearing Aids	78.0	77.9	80.9	CD
Larger Ears	76.0	81.4	84.1	B*
Bookshelf Speakers	75.7	76.6	77.3	A
Hi Rated Phones	75.5	62.0	72.5	A
2:1 Compression	72.7	83.1	89.0	B*
OTE-40	62.5	50.7	62.1	A
Popular Phones	62.0	43.7	58.1	CD
Airline Phones	60.0	53.3	51.4	A
Discount Stereo	47.8	45.3	59.9	CD
Speech Audiometer	46.9	42.0	53.8	CD
Monophonic	43.4	20.0	33.8	B
Pocket Radio	16.0	18.7	13.4	CDA
Estimated Standard Error	2.5%	3.6%	2.5%	

* Note: Only two ratings (orchestra and piano trio) available on tape B. Inclusion of a Live-Voice rating would probably have increased the average rating of the first three of these systems.

accomplished on a manikin than a live subject) can drop the ratings into the 80 percent region. This is true even though the "Larger Ears" were intentionally designed to produce nearly the same frequency response as the original KEMAR ears (Maxwell and Burkhard, 1978).

Differences among commonly sold high-fidelity systems can produce ratings in the 60 to 80 percent region. The 78 to 90 percent ratings obtained by the various experimental hearing aids clearly place them in what is commonly considered the "high-fidelity" class.

The 84- to 96-percent rating given the "+3 dB" comparisons by the three subject groups indicates two things. Even when people try to ignore level differences they are not completely successful; but the effect of small errors in setting comparison levels was probably not important in these experiments.

With regard to hearing-aid design, several things seem clear. First of all, undamped earmold-tubing resonances and failure to take head diffraction and external-ear resonances into account will not result in a bad rating for a hearing aid as long as no other defects (distortion, poor low-frequency response, etc.) exist simultaneously, as indicated by the ratings for the OTE-40 aids. Indeed, these aids generally rated higher than the "Airline Phones", whose one-third-octave frequency response was similar. As a corollary, the common complaint regarding hearing aid sound quality may be related more to amplifier distortion due to peak clipping than to the frequency response of the hearing aid per se (a conclusion known for some time by knowledgeable hearing-aid designers). Secondly, while an In-The-Concha microphone location may be required to produce nearly perfect

fidelity ratings, even an Over-The-Ear microphone location can produce very good ratings in a properly designed hearing aid. This is consistent with the conclusions of the informal pinna experiments discussed in Chapter II. Thirdly, the highest fidelity (similarity) rating given any system was given by all three subject groups to the "Changed Microphones" system. This indicates that the improved transient response which presumably results from reducing the "ringing" near 8 kHz (because of what, in electrical filter theory, can be humorously referred to as a "pile of poles" near 8 kHz when both the earphone and microphone have a sharp cutoff at 8 kHz) produces very little audible difference in sound quality.

A determination of the optimum form of signal processing for a high-fidelity hearing aid was beyond the scope of the present study, but the 73- to 89-percent fidelity ratings given the "2:1 compression" condition by the three subject groups indicates that such compression is an acceptable form of signal processing. Indeed, the Trained-Listener subjects, who were the most successful in following the instruction to ignore loudness differences, gave an average rating of 89 percent to the 2:1 compression condition. (The gain introduced by the compression amplifiers acted to increase the loudness of all except the most intense portions of the program selections.) It would thus appear that there is no need to wait until the optimum form of signal processing has been established before truly high-fidelity hearing aids can be produced. Compression amplification (in this case having input-output characteristics essentially similar to those shown in Figure 2-12 on page 78) can be employed without serious damage to

sound quality, a conclusion consistent with the widely accepted use
of such signal processing in the broadcast and recording industries.

The "monophonic" condition obtained the next-to-lowest rating
(only the $4.95 Pocket Radio obtained lower ratings) from each subject
group. In the case of the Trained-Listener subjects and Golden-Ear
subjects, the monophonic condition meant that only one (the left) ear
received stimulation during the "B" portion of the ABA comparison.
In the case of the Untrained-Listener subjects, only the left loud-
speaker was heard during the B portion. In the former case, the
fidelity rating fell in the Bad-to-Very-Bad category. In the latter
case, the fidelity rating fell in the Fair-to-Bad category. Both
results appear to be more severe than Fletcher's (1942) findings that
subjects would prefer two-channel reproduction with a 5-kHz cutoff
frequency to one-channel reproduction with unlimited bandwidth. In
any case, they stand in amusing contrast to the recent FTC order that
binaural hearing aids cannot be advertised as generally beneficial.

D. Accuracy Scores versus Fidelity Ratings

A comparison between objective and subjective measures of sound-
system performance is possible by comparing the 21-band accuracy score
calculations (applied to the relative response of each system) to the
subjective fidelity ratings. A plot of the fidelity ratings versus
21-band accuracy scores for 12 of the systems whose one-third-octave
frequency response was known revealed good correlation in all except
three cases: one (pocket radio) in which severe nonlinear distortion
was combined with response inaccuracies, and two in which the low-bass

response was exaggerated (system PP) or completely missing (system SA).
Thus an extension of the 21-band accuracy score calculations to include
four more bands in the bass region (center frequencies of 50, 63, 80,
and 100 Hz) was performed to produce a "25-band" accuracy score.
Table VII contains the 21-band and 25-band accuracy scores and the
subjective ratings for the 12 systems.

Figure 4-9 shows the data of Table VII as a comparison between the
subjective fidelity ratings obtained from each subject group and the
calculated 25-band accuracy scores. The correlation coefficients were
$r = .89$ (Untrained Listeners), $r = .93$ (Golden Ears), and $r = .95$
(Trained Listeners), respectively. Although there were fairly large
discrepancies in individual cases, these correlations indicate that
the 25-band accuracy score would be a simple and us, ;ul first step in
estimating the subjective fidelity of a sound-reproduction system.

For our present purposes, note that the 12 systems can be con-
veniently grouped into three regions: a "good-to-excellent" fidelity
(similarity) region, a "fair" fidelity region, and a "bad" region.
Of importance here is the fact that all three of the experimental
hearing aid pairs obtained fidelity ratings and 25-band accuracy
scores in the "good-to-excellent" region, as did the more accurate
loudspeakers and headphones.

It is well at this point to recall again that these accuracy
scores were applied to the relative response of each system as shown
in Figure 4-2, just as the fidelity ratings were similarity ratings
between the sound of the comparison and reference systems. Thus a
crucial assumption underlying the interpretation of the results of

TABLE VII

TWENTY-ONE-BAND AND TWENTY-FIVE-BAND ACCURACY SCORES AND SUBJECTIVE
FIDELITY RATINGS (AVERAGED OVER THREE SOURCE PROGRAMS)
FOR TWELVE SOUND SYSTEMS

Sound System	21-Band Accuracy Score	25-Band Accuracy Score	Average Fidelity Ratings		
			Untrained Listeners	Trained Listeners	Golden Ears
ITE	91	90	78	90	78
ITC	92	87	80	90	88
BS	88	86	76	77	77
OTE	82	82	78	81	78
MS	83	81	80	86	73
RR	86	78	84	87	78
HRP	76	74	76	73	62
SA	79	69	47	54	42
PP	75	61	62	58	44
OTE-40	54	59	63	62	51
AP	62	56	60	51	53
PR	38	32	13	13	19

Note: Listed in order of 25-Band Accuracy Score
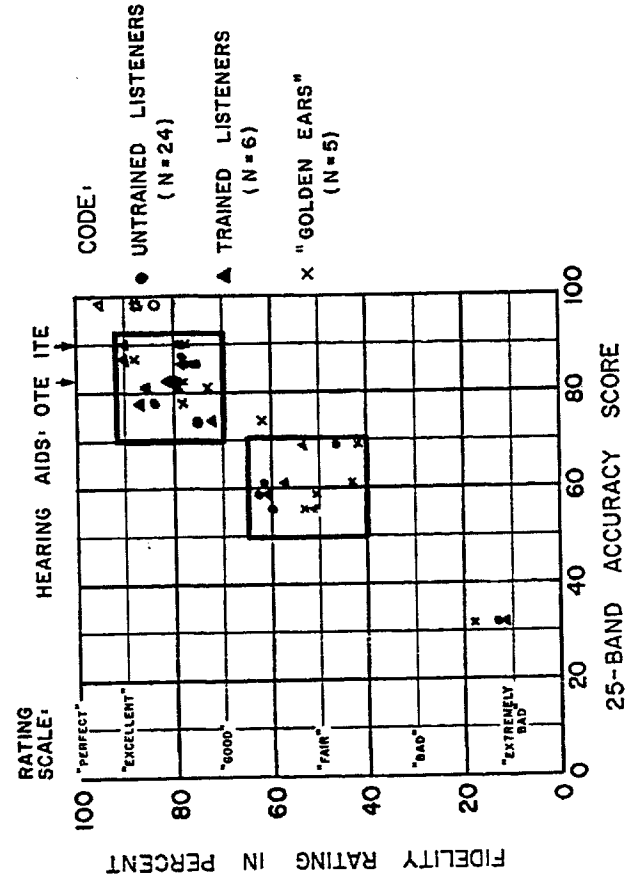
FIGURE 4-7 COMPARISON BETWEEN AVERAGE SUBJECTIVE FIDELITY (SIMILARITY) RATINGS AND CALCULATED 25 BAND ACCURACY SCORES FOR 12 SYSTEMS RATED ON THREE IDENTICAL PROGRAM SELECTIONS (ORCHESTRA, PIANO TRIO, MALE VOICE). OPEN SYMBOLS OBTAINED FROM "+3 dB" LEVEL SHIFT.

these experiments was that the reference system itself had good fidel-
ity. This question was addressed in a final experiment discussed
below.

E. Live-Versus-Recorded Fidelity Rating of Reference System

As mentioned above, a true live-voice "live-versus-recorded" set
of comparisons was included on comparison tape B. This turned out to
be a difficult challenge in itself, since informal auditing of trial
comparisons indicated that the position of the talker relative to the
loudspeaker was extremely critical. Thus, one comparison included
on tape B was between the writer as talker ("Joe took father's...")
standing first beside the loudspeaker and then leaning back roughly
.5 meter into the corner of the recording room. The across-subject-
group's fidelity rating for this comparison was only 60 percent!
Several factors presumably account for that low rating. The increased
"bass response" obtained from corner sources has already been noted.
In addition, given the size of the head, virtual images due to corner
reflections (Knowles, 1941) should have produced interference effects
in the 1- to 2-kHz region, a region critical to voice-quality judgments.
Finally, the writer's head-to-torso relationship was unavoidably warped
by the "leaning back into the corner" process.

The across-subject-group rating obtained for a live-voice compari-
son between the talker standing immediately beside the left AR3a
monitor speaker and that loudspeaker reproducing a previous anechoic-
chamber recording of the talker was a relatively high 80 percent. The
comparable rating for the monitor loudspeakers, however, was only
53 percent.

These somewhat puzzling results were discussed with Edgar Villchur,
the designer of the original AR3 loudspeakers and a designer with ex-
tensive experience in live-versus-recorded comparisons. His explanation
appears to be the obvious one. The human voice is highly directional.
An anechoic chamber recording of a talker provides an on-axis recording
of his voice whose spectral analysis will differ appreciably from the
far-field "power response" of his voice as measured in a moderately
reverberant room. In particular, the on-axis response of the voice will
sound "bright" (have apparent excessive high-frequency energy) compared
to its power response.

Indeed, careful auditing of the comparison in question revealed
that the AR3a reference speaker appeared to have slightly too much
high-frequency response, despite the high-frequency rolloff in the
one-third-octave room response of the AR3a's shown in Figure 4-1!
This problem was exaggerated by the relatively flat high-frequency
room response of the Monitor Speaker system. It is clear that achiev-
ing true fidelity is a tricky business. It is also clear that the
fidelity--i.e., similarity to the reference loudspeaker--ratings for
the monitor speakers would have been significantly higher if their
high-frequency level control had been adjusted to produce the best
match to the reference speakers rather than to the manufacturer's
"flat response" setting. Indeed, Schulein (1975) found the as-adjusted
room response of monitor loudspeakers in recording studios had even
more high-frequency rolloff than shown in Figure 4-1.

In any case, the choice of reference loudspeakers appeared to have
been adequate to the present purposes.

Finally, the right AR3a supertweeter failure must be addressed. The relatively high rating given the OTE aids with 8 kHz cutoff was of some comfort, since the supertweeter came into play only above 8 kHz. The change in rating of system MS against the repaired versus unrepaired reference systems (in the "extra" orchestral comparison) was -1.4 percent (Untrained Listeners), +4 percent (Golden Ears), and +6.3 percent (Trained Listeners). None of these achieved statistical significance, but it appears likely that the overall rating for those systems with "flat" high-frequency response would have been a few percent higher if the tweeter failure had not occurred.

All things considered, it seems reasonable to conclude that the present fidelity-rating experiments had both good reliability and good validity. That is, essentially similar fidelity ratings would undoubtedly be obtained if the experiment were to be repeated without flaw, and the fidelity ratings obtained here probably provide a good estimate of the ratings which might be assigned in real-life situations.

CHAPTER V

SUMMARY AND CONCLUSIONS

Experimental hearing aids designed to meet a priori guidelines for high-fidelity performance were rated by three subject groups on a simulated live-versus-recorded comparison task, along with several common high-fidelity loudspeakers and headphones. The fidelity ratings given the hearing aids by all three subject groups placed them in the Good-to-Excellent fidelity class. The hearing-aid ratings were similar to those given to very-high-quality sound systems.

Several conclusions flow from these results:

1. The design guidelines given in Chapter II appear adequate to insure a high-fidelity rating for a completed hearing aid.

2. An exaggerated bass and treble response may sell well in hi-fi dealer's show rooms, but does not produce good fidelity ratings.

3. One should use the phrase "high-fidelity speech audiometer" with discretion.

4. As suggested some decades ago by Fletcher, an 8 kHz bandwidth is sufficient to produce "substantially complete fidelity".

5. Two ears are better than one. As a corollary, two hearing aids will undoubtedly be better than one for hearing-impaired users with good binaural hearing.

6. While an In-The-Concha microphone location may be required to produce nearly perfect fidelity ratings, even an Over-The-Ear microphone location can produce very good ratings in a

properly designed hearing aid.

7. Undamped earmold tubing resonances and failure to take head diffraction and external-ear resonances into account will not result in a bad rating for a hearing aid as long as no other defects (distortion, poor low-frequency response, etc.) exist simultaneously. As a corollary, the common complaint regarding hearing aid sound quality may be related more to amplifier distortion due to peak clipping than to the frequency response of the hearing aid per se.

8. Untrained listeners appear to produce ratings similar to those produced by trained listeners, if allowance for differences in overall levels are made.

9. Trained listeners produce greatly reduced variability, so that one trained listener can sometimes be worth many untrained listeners in sample size considerations.

The most important conclusion reached in this study is that current hearing aid amplifier and transducer technology does, in fact, permit the construction of practical high-fidelity hearing aids as judged by those with normal hearing. Not surprisingly perhaps, at least one high-fidelity hearing aid design was being made commercially available shortly after the present study was undertaken (Toepholm, 1979). A pair of these aids could not be included in the present fidelity-rating experiment, but the writer has conducted recent informal listening-test evaluations which indicate the "high fidelity" era in hearing aid design has already begun.

A. Future Research: Hearing Aids

There are at least three reasons for demonstrating that it is possible to design a hearing aid which is judged high-fidelity by someone with _normal_ hearing:

1. Such a design provides a convenient base to which electronic signal processing can conveniently be added;

2. A hearing aid which provides gain only for low-level signals (i.e., is a unity-gain high-fidelity sound-reproduction system for high-level signals) may prove useful to a large number of individuals; and most importantly,

3. The demonstration supports the following conclusions: The important question for hearing aid research is no longer "What _can_ a hearing aid be designed to do?", but "What _should_ a hearing aid be designed to do for the hearing impaired?".

The lack of a satisfactory answer to the last question is a major barrier to vastly improved hearing aid design. That question can be restated: What hearing aid characteristics will prove to be optimum (or even somewhere near optimum) for a given individual as he goes about his daily life? As a specific example; will a substantial number of hearing aid users with mild-to-moderate hearing impairments prefer a high-fidelity hearing aid (as defined in Chapter II, Section I-C) to a more conventional hearing aid? More specifically, what type of automatic gain control will most unobtrusively provide the variable-gain (and variable-frequency-response) amplification presumably required of a high-fidelity _hearing aid_?

Preliminary answers to these questions could be obtained in laboratory experiments such as the fidelity-rating experiment described in Chapter IV, but one suspects that the final answers can only be obtained through the fairly clumsy process of trial and error in the marketplace, as dispensers discover which new hearing aid designs provide increased user satisfaction.

B. Future Research: High Fidelity

Several fidelity-related experiments suggested themselves during the course of the present investigation. The writer would have liked to participate in a commercial recording session--of the Chicago Symphony Orchestra, for example--with a KEMAR manikin in a "representative" audience seat (several seats would allow a better experiment) during a live performance. A comparison of that "live" eardrum-pressure recording with an eardrum-pressure recording made during a subsequent reproduction of the resulting 33-1/3 rpm commercial disc would allow a true test of the overall fidelity of the current recording-reproduction process.

A simpler but related experiment would be useful: simply comparing the long-term spectra obtained from an equalized-manikin recording of several live performances, with the long-term spectra subsequently measured on commercial discs of the same performances, would provide useful objective information regarding the appropriate room-response of high-fidelity loudspeakers (and the real-ear response of high-fidelity headphones).

REFERENCES

Ahren, T., Arlinger, S., Holmgren, C., Jerlvall, L., Johnsson, B., Lindblad, A-C., Persson, L., Pettersson, A. and Sjögren, G. (1977). "Automatic gain control and hearing aids," Report TA No. 84, Karolinska Institutet, Tech. Audiol. (Stockholm).

Allison, R.F. and Berkovitz, R. (1972). "The sound field in home listening rooms," J. Audio Eng. Soc. 20, 459-469.

Anon (1977). "How CU's audio lab tests loudspeaker accuracy," Consumer Reports TNG-3 (Mount Vernon, N.Y.).

ANSI (1969). "Specifications for audiometers," S3.6-1969 (Am. Natl. Stand. Inst., New York).

Barfod, J. (1972). "Investigations on the optimum corrective frequency response for high tone hearing loss," Report No. 4, Acoustics Laboratory, Technical Univ. of Denmark.

Barfod, J. (1976). "Multi-Channel Compression Hearing Aids," Report No. 11, Acoustics Laboratory, Tech. Univ. of Denmark.

Barfod, J. (1978). "Multichannel compression hearing aids: experiments and consideration on clinical applicability," in Sensorineural hearing impairment and hearing aids (Ludvigsen, C. and Barfod, J. eds.) Scand. Audiol. Suppl. 6, 315-340.

Bauer, B.B. (1945). "Frequency response preference tests for phonograph reproduction," J. Acoust. Soc. Am. 17, 102(A).

Bauer, B.B. (1974). "Audibility of phase distortion," Wireless World, (March, 1974), 27-28.

Bekesy, G. von (1934). "Über die nichtlinearen verzerrungen des ohres," Ann. Physik, 20, 809-827. (Also, pp. 332-344 in Experiments in Hearing; McGraw-Hill; N.Y.: 1960).

Berland, O. and Nielsen, E. (1969). "Sound pressure generated in the human external ear by a free sound field," Audecibel (Summer 1969) 103-109.

Berlin, C.I. and Gondra, M.I. (1976). "Extratympanic clinical cochleography with clicks," Electrocochleography (eds. Rubin, Eberling and Salomon; Univ. Park Press, Baltimore).

Bevin, W.R. (1978). "A simplified equalization analyzer," J. Audio Eng. Soc. 26, 120-129.

Blesser, B.A. and Ives, F. (1972). "A reexamination of the S/N question for systems with time-varying gain or frequency response," J. Audio Eng. Sc 20, 638-641.

Briskey, R.J. (1976). "The importance of SSPL," Beltone Colloquy No. 1 (June, 1976), (Beltone, Chicago).

Brownlee, R.G. (1965). Statistical Theory and Methodology in Science and Engineering, (Wiley, N.Y.).

Bucklein, R. (1962). "Horbarkeit von Unregelmakigkeiten in Frequenzgangen bei akustischer Ubertragung," Frequenz 16, 103-108.

Burkhard, M.D. (1978). Manikin Measurements--Conference Proceedings, edited by M.D. Burkhard (Industrial Research Products, Elk Grove Village, Ill.).

Burkhard, M.D. and Sachs, R.M. (1975). "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. 58, 214-222.

Burnett, E.D., Corliss, E.L.R., and Berendt, R.D. (1972). "Magnetic recording of acoustic data on audiofrequency tape," Nat. Bur. Stand. (U.S.) Tech. Note 718.

Butler, R.A. and Belendiuk, K. (1977). "Spectral cues utilized in the localization of sound in the median saggital plane," J. Acoust. Soc. Am. 61, 1264-1269.

Carhart, R. (1946). "Tests for selection of hearing aids," Laryngoscope 56, 780-794.

Carhart, R. (1950). "Hearing Aid Selection by University Clinics," J. Speech Hear. Disorders 15, 106-113.

Carlson, E.V. (1974). "Smoothing the hearing aid frequency response," J. Audio Eng. Soc. 22, 426-429.

Carlson, E.V., Mostardo, A.F. and Diblick, A.V. (1976). "Transducer with improved armature and yoke construction," U.S. Patent # 3,935,398 (filed July, 1971).

Ceoen, C. (1972). "Comparative stereophonic listening tests," J. Audio Eng. Soc. 20, 19-27.

Clack, T.D., Erdreich, J. and Knighton, R.W. (1972). "Aural Harmonics: The monaural Phase Effects," J. Acoust. Soc. Am. 52, 536-541.

Consumer Reports (1977). "Low-priced loudspeakers," Consumer Reports 42, 406-409.

Consumer Reports (1978). "High-priced loudspeakers," Consumer Reports 43, 592-599.

Corliss, E.L.R., Kobal, M. and Berghorn, S.E. (1960). "Reduction in signal-to-noise ratio induced by distortion in speech transmission systems," J. Acoust. Soc. Am. 32, 1502(A).

Cox, R.M. and Studebaker, G.A. (1979). "Problems in the recording and reproduction of hearing aid processed signals," in Acoustical Factors Affecting Hearing Aid Performance and Measurement, G. Studebaker and I. Hochberg, eds. (University Park Press, Baltimore).

Dallos, P., Ryan, A., Harris, D., McGee, T. and Özdamar, O. (1977). "Cochlear frequency selectivity in the presence of hair cell damage," in Psychophysics and Physiology of Hearing, Evans, E.F. and Wilson, J.P. (eds.) (Academic, London).

Dalsgaard, S.C. and Jensen, O.D. (1974). "Measurement of insertion gain of hearing aids," Eighth Int. Cong. on Acoust. (London) Vol. 1, 205.

Davies, O.L. (ed.), Box, G.E.P., Conner, L.R., Cousins, W.R., Itimsworth, F.R. and Sillitto, G.P. (1960). Design and Analysis of Industrial Experiments (Hafner, N.Y.).

Davis, H., Hudgins, C.V., Peterson, G.E. and Ross, D.A. (1946). "Desirable frequency characteristics for hearing aids," J. Acoust. Soc. Am. 18, 247(A).

Davis, H., Stevens, S.S., Nichols, R.H., Jr., Hudgins, C.V., Marquis, R.J., Peterson, G.E. and Ross, D.A. (1947). Hearing Aids, An Experimental Study of Design Objectives (Harvard Univ. Press, Cambridge).

Davis, M. (1978). "What's really important in loudspeaker performance?," High Fidelity (June, 1978), 53-58.

deBoer, E. and Bouwmeester, J. (1975). "Clinical Psychophysics," Audiology 14, 274-299.

Dodds, E. and Harford, E. (1968). "Modified earpieces and CROS for high frequency hearing losses," J. Speech Hear. Res. 11, 204-228.

Dunn, H.K. and Farnsworth, D.W. (1939). "Exploration of pressure field around the human head during speech," J. Acoust. Soc. Am. 10, 184-199.

Dunn, H.K. and White, S.D. (1940). "Statistical measurements on conversational speech," J. Acoust. Soc. Am. 11, 278-288.

Egan, J.P. and Klumpf, R.C. (1951). "The error due to masking in the measurement of aural harmonics by the method of best beats," J. Acoust. Soc. Am. 23, 275-286.

Evans, E.F. (1978). "Peripheral auditory processing in normal and abnormal ears: physiological considerations for attempts to compensate for auditory deficits by acoustic and electrical prosthesis," in Sensorineural hearing impairment and hearing aids (Ludvigseu, C., and Barfod, J., eds.) Scand. Audiol. Suppl. 6, 9-48.

Filler, A.S., Ross, D.A. and Wiener, F.M. (1945). "The pressure distribution in the auditory canal in a progressive sound field," Harvard University Report PNR-5 (Contract N5ori-76) to Office of Research and Inventions, U.S. Navy.

Flanagan, J. (1957). "Difference limen for formant amplitude," J. Speech Hear. Dis. 22, 205-212.

Fletcher, H. (1931). "Some physical characteristics of speech and music," Bell Syst. Tech. J. 10, 349-373.

Fletcher, H. (1942). "Hearing, the determining factor for high-fidelity transmission," Proc. Inst. Rad. Eng. 30, 266-277.

Fletcher, H. (1953). Speech and Hearing in Communication (Van Nostrand, reprinted Krieger, N.Y., 1972).

Fournier, J.E. (1965). "La Correction Prothètique des Studités: Problemes Techniques a Medico-Sociaux," Practk. oto-rhino-laryng. 27, 180-198.

Fournier, J.E. and Rainville, M.,-J. (1967). "Le Controle des Resultats D'Appareillage par Audiometrie Automatique en Champ Libre," Int. Audiol. 6, 264-267.

French, N.R. and Steinberg, J.C. (1947). "Factors governing the intelligibility of speech," J. Acoust. Soc. Amer. 19, 90-119.

Gabrielsson, A., Rosenberg, U. and Sjögren, H. (1974). "Judgments and dimension analyses of perceived sound quality of sound-reproducing systems," J. Acoust. Soc. Am. 55, 854-861.

Gabrielsson, A. and Sjögren, H. (1976). "Preferred listening levels and perceived sound quality at different sound levels in 'high fidelity' sound reproduction, Report TA No. 82, Karolinska Institutet, Technical Audiology (Stockholm), 1-33 plus appendices.

Gabrielsson, A., Nyberg, P.O., Sjögren, H. and Svensson, L. (1976). "Detection of amplitude distortion by normal hearing and hearing impaired subjects," Report TA No. 83, Karolinska Institutet, Technical Audiology (Stockholm).

Gannett, D.K. and Kerney, I. (1944). "The discernibility of changes in program band width," Bell Syst. Tech. J. 23, 1-10.

Gengel, R.W. (1973). "Temporal effects in frequency discrimination by hearing impaired listeners," J. Acoust. Soc. Am. 54, 11-15.

Goldberg, H. (1960). "A new concept in hearing aids," Dynaura, Inc. Report (Flushing, N.Y.).

Goldberg, H. (1965a). "Psychological Adaptation to Hearing Correction," Analog of MPLS dealer bulletin.

Goldberg, H. (1965b). "Hearing aids and the profoundly deaf," The Hearing Dealer (May, 1965), 13-15.

Goldberg, H. (1966). "Hearing Aid," U.S. Patent No. 3229049 (Filed Aug. 4, 1960).

Goldberg, H. (1972). "The utopian hearing aid: Current state of the art," J. Aud. Res. 12, 331-335.

Goldstein, J.L. (1967). "Auditory nonlinearity," J. Acoust. Soc. Am. 41, 676-689.

Hall, J.L. (1972). "Auditory distortion products $f_2 - f_1$," J. Acoust. Soc. Am. 51, 1863-1871.

Halperin, H.R., Cullon, J.K., Berlin, C.I. and Killion, M.C. (1977). "Translating Hearing Aid for 'Deaf' Patients with Nearly Normal Ultra-Audiometric Hearing," J. Acoust. Soc. Am. 62, 576(A).

Harris, J.D. (1971). "The true goals and aspirations of audiology," in Harris on Audiology (1973) (J. Auditory Res., Groton).

Helle, R. (1969). "Amplitude und Phase des im Gehor gebildefen differenztones dritter ordung," Acoustica 22, 74-87.

Hood, J.D. (1977). "Loudness Balance procedures for the measurement of recruitment," Audiology 16, 215-228.

Hoth, D.F. (1941). "Room noise spectra at subscriber's telephone locations," J. Acoust. Soc. Am. 12, 499-504.

Ingelstam, R., Johansson, B., Pettersson, A. and Sjögren, H. (1971). "The effect of non-linear amplitude distortion, an investigation by variation of the quadratic and the cubic components," Report TA No. 64, Karolinska Institutet, Tech. Audiology (Stockholm), 1-5.

Jerger, J.F., Tillman, T.W. and Peterson, J.L. (1960). "Masking by octave bands of noise in normal and impaired ears," J. Acoust. Soc. Am. 32, 385-390.

Johansson, B., Sjögren, H. and Nyman, H. (1974). "The effect of amplitude distortion in hearing aids on the intelligibility of speech," Eighth Int. Cong. Acoust. (London), 209.

Katz, D.R. and Longinotti, C. (1977). "Speech discrimination scores in hearing aid selection procedure," Presented at November, 1977 Amer. Speech Hear. Assoc. convention.

Killion, M.C. (1976a). "Noise of Ears and Microphones," J. Acoust. Soc. Am. 59, 424-433.

Killion, M.C. (1976b). "Earmold Plumbing for Wideband Hearing Aids," J. Acoust. Soc. Am. 59, 562(A). Available from Knowles Electronics, Franklin Park, IL.

Killion, M.C. (1978). "Revised estimate of minimum audible pressure: Where is the 'missing 6 dB'?," J. Acoust. Soc. Am. 63, 1501-1508.

Killion, M.C. (1979a). "Equalization filter for eardrum-pressure recording using a KEMAR manikin," J. Audio Eng. Soc. 27, 13-16.

Killion, M.C. (1979b). "Problems in the application of wideband hearing aid earphones," in Acoustical Factors Affecting Hearing Aid Measurement and Performance, G.A. Studebaker and I. Hochberg, eds. (University Park Press, Baltimore).

Killion, M.C. and Carlson, E.V. (1970). "A wideband miniature microphone," J. Audio Eng. Soc. 18, 631-635.

Killion, M.C., Carlson, E.V. and Burkhard, M.D. (1970). "Audio frequency amplification circuit," U.S. Patent No. 3,512,100 (Filed 1968).

Killion, M.C. and Carlson, E.V. (1974). "A subminiature electret-condenser microphone of new design," J. Audio Eng. Soc. 22, 237-243.

Killion, M.C. and Monser, E.L. (1979). "CORFIG: COupler Response for Flat Insertion Gain," in Acoustical Factors Affecting Hearing Aid Performance, G.A. Studebaker and I. Hochberg, eds. (University Park Press, Baltimore).

Killion, M.C. and Studebaker, G.A. (1978). "A-weighted equivalents of permissible ambient noise during audiometric testing," J. Acoust. Soc. Am. 63, 1633-1635.

Knowles, H.S. (1941). "Loud-speakers and room acoustics," in Radio Engineering Handbook, K. Henney, ed. (McGraw-Hill, New York), 876-928.

Knowles, H.S. (1959). "Physical aspects of hearing aids," Presented at first HAIL workshop, Allerton House, Univ. of Ill.

Knowles, H.S. (1967). "Some problems in headworn aid response measurements," presented at the 33rd convention, Audio Eng. Soc. (Oct. 16).

Knowles, H.S. (1968). Comments on MEDRESCO hearing aids - personal communication.

Knowles, H.S. and Killion, M.C. (1978). "Frequency Characteristics of Recent Broadband Receivers," J. Audio Tech. (Zeitschrift für Hörgerate-Akustik) 17, 86-99 and 136-140.

Kohler, I. (1962). "Experiments with goggles," Scientific Amer. (May, 1962).

Kuhn, G.F. and Burnett, E.D. (1977). "Acoustic pressure field alongside a manikin's head with a view towards in situ hearing-aid tests," J. Acoust. Soc. Am. 62, 416-423.

LeBel, C.J. (1944). "Pressure and field response of the ear in hearing aid performance determination," J. Acoust. Soc. Am. 16, 63-67.

Licklider, J.C.R. (1949). "Basic correlates of the auditory stimulus," in Handbook of Experimental Psychology, S.S. Stevens, ed. (Wiley, N.J.).

Licklider, J.C.R. and Pollack, I. (1948). "Effects of differentiation, integration, and infinite peak clipping on the intelligibility of speech," J. Acoust. Soc. Am. 20, 42-51.

Lippman, R.P. (1978). "The effect of amplitude compression on the intelligibility of speech for persons with sensorineural hearing loss," Ph.D. Thesis, Mass. Inst. Tech. (University Microfilms, Ann Arbor).

Lybarger, S.F. (1944). U.S. Patent Application S.N. 543,278, July 3, 1944.

Lybarger, S.F. (1972). "Earmolds," in Handbook of Clinical Audiology, J. Katz, Ed. (Williams and Wilkins, Baltimore).

Lybarger, S.F. (1974). "The acoustical engineer's viewpoint of hearing aid design," Bul. N.Y. Academy Med. 50, 917-930.

Lybarger, S.F. (1978). "Selective Amplification--a review and evaluation," J. Amer. Audio. Soc. 3, 258-266.

Madaffari, P.L. (1974). "Pressure variation about the ear," J. Acoust. Soc. Am. 56, S3(A) Available from Knowles Electronics, Franklin Park, Il.

Marsh, R.C. (1975). "Tweeters in the grass, alas," Chicago (Aug., 1975), 76-78.

Martin, M.C. (1971). "Electroacoustic characteristics of hearing aids and sensorineural hearing loss," in Electroacoustic characteristics relevant to hearing aids, H.W. Ewertsen, ed. Scand. Audiol. Suppl. 1, 105-108.

Martin, M.C. (1973). "Hearing aid gain requirements in sensorineural hearing loss," British J. Audiol. 7, 21-24.

Mathes, R.C. and Wright, S.B. (1934). "The compandor--an aid against static in radio telephony," Bell Syst. Tech. J. 13, 315-322.

Maxwell, R. and Burkhard, M.D. (1978). "Larger ear replica for KEMAR manikin," Indust. Res. Products Rep. No. 20059-1 to Knowles Electronics (Franklin Park, IL).

McGee, T. (1978a). "Psychophysical tuning curves from hearing impaired listeners," Ph.D. dissertation, Northwestern Univ. (University Microfilms, Ann Arbor).

McGee, T. (1978b). Clinical Observations on Brain-Stem Evoked Response (personal communication).

Miller, J.D. (1974). "Effects of noise on people," J. Acoust. Soc. Am. 56, 729-764.

Miller, J.D. and Niemoeller, A.F. (1967). "Hearing aid design and evaluation for a patient with a severe discrimination loss for speech", J. Speech Hearing Res. 10, 367-372.

Miller, R. (1966). Simultaneous Statistical Inference (McGraw-Hill, N.Y.)

Milner, P. (1977). "How much distortion can you hear?," Stereo Rev. (June, 1977), 64-68.

Moe, C.R. (1942). "An experimental study of subjective tones produced within the human ear," J. Acoust. Soc. Am. 13, 159-166.

Moir, J. (1958). High Quality Sound Reproduction (Macmillan, N.Y.) p. 51.

Møller, H. (1974). "Relevant loudspeaker tests in studios, in Hi-Fi dealers' demo rooms, in the home, etc.--using 1/3 octave, pink-weighted, random noise," presented at the 47th Audio Eng. Soc. Convent. (Copenhagen). Available as B&K application note No. 17-197.

Mott, E.E. (1944). "Indicial Response of Telephone Receivers," Bell System Tech. Jour. 23, 135-150.

Munson, W.A. and Wiener, F.M. (1952). "In search of the missing 6 dB," J. Acoust. Soc. Am. 24, 498-501.

Nixon, G.M. (1945). "'Higher Fidelity' in Sound Transmission and Reproduction," J. Acoust. Soc. Am. 17, 132-135.

Olson, H.F. (1957). Acoustical Engineering (Van Nostrand, N.Y.).

Olson, H.F. (1967). Music, Physics, and Engineering (Dover, N.Y.).

Palva, T., Goodman, A. and Hirsh, I.J. (1953). "Critical evaluation of noise audiometry," Laryngoscope 63, 842-860.

Pascoe, D.P. (1975). "Frequency response of hearing aids and their effects on the speech perception of hearing-impaired subjects," Annals Otology, Rhin., Laryng. 84, Supp. 23.

Peters, R.W. and Burkhard, M.D. (1968). "On noise distortion and harmonic distortion measurements," Industrial Res. Prod. Report No. 10350-1 to Knowles Electronics (Franklin Park, IL.).

Peterson, A.P.G. (1951). "Intermodulation distortion," General Radio Experimenter 25, No. 10.

Punch, J.L. (1978). "Quality judgments of hearing aid-processed speech and music by normal and otopathologic listeners," J. Amer. Audiol. Soc. 3, 179-188.

Rabinowitz, W.M. (1976). "Effects of the acoustic reflex on middle-ear transmission and ear-drum admittance in man," J. Acoust. Soc. Am. 59 Suppl. No. 1, S29(A).

Radley, W.G., Bragg, W.L., Dadson, R.S., Hallpike, C.S., McMillan, D., Pocock, L.C. and Littler, T.S. (1947). "Hearing Aids and Audio-meters," Report of the Committee on Electroacoustics of the Medical Research Council (His Majesty's Stationery Office, London), 1-71.

Rhodes, L.M. (1969). "What you should know about hearing aids," Today's Health (Aug., 1969).

Romanow, F.F. (1942). "Methods for measuring the performance of hearing aids," J. Acoust. Soc. Am. 13, 294-304.

Rosenthal, R. (1975). The Hearing Loss Handbook (St. Martin's, N.Y.).

Sachs, R.M. and Burkhard, M.D. (1972). "Earphone pressure response in ears and couplers," J. Acoust. Soc. Am. 51, 140(A). Available as Industrial Research Products Report No. 20021-2 to Knowles Electronics.

Scharf, B. (1978). "Comparison of Normal and Impaired Hearing. I. Loudness, Localization: II. Frequency Analysis, speech perception," in Sensorineural hearing impairment and hearing aids (eds. Ludvigsen, C. and Barfod, J.), Scand. Audiol. Suppl. 6, 4-106.

Scharf, B. and Hellman, R.P. (1966). "A model of loudness summation applied to impaired ears," J. Acoust. Soc. Am. 40, 71-78.

Schuknecht, H.F. (1974). Pathology of the Ear (Harvard Press, Cambridge).

Schulein, R.B. (1975). "In Situ measurement and equalization of sound reproduction systems," J. Audio Eng. Soc. 23, 178-186.

Schweitzer, H.C. and Causey, G.D. (1977). "The relative importance of recovery time in compression hearing aids," Audiology 16, 61-72.

Seacord, D.F. (1940). "Room noise at subscriber's telephone locations," J. Acoust. Soc. Am. 12, 183-187.

Shaw, E.A.G. (1966). "Earcanal pressure generated by circumaural and supra-aural earphones," J. Acoust. Soc. Am. 39, 471-479.

Shaw, E.A.G. (1974). "The external ear," in Handbook of Sensory Physiology, W.D. Keidel and W.D. Neff, Eds. (Springer-Verlag, Berlin), Vol. VII, Chapter 14.

Shaw, E.A.G. (1976). "Diffuse field sensitivity of the external ear based on the reciprocity principle," J. Acoust. Soc. Am. 60, S102(A).

Shore, I., Bilger, T. and Hirsh, I.J. (1960). "Hearing aid evaluation: reliability of repeat measurements," J. Speech Hear. Dis. 25, 152-170.

Silverman, S.R. and Taylor, S.G. (1947). "The choice and use of hearing aids," in Hearing and Deafness (Rinehart, N.Y.).

Skinner, M.W. (1976). "Speech intelligibility in noise-induced hearing loss: Effects of high-frequency compensation," Ph.D. Thesis, Washington University (University Microfilms, Ann Arbor).

Snow, W.B. (1931). "Audible frequency ranges of music, speech, and noise," Bell Sys. Tech. J. 10, 616-627.

Speith, W. (1957). "Downward spread of masking," J. Acoust. Soc. Am. 29, 502-505.

Steinberg, J.C. and Gardner, M.B. (1937). "The dependence of hearing impairment on sound intensity," J. Acoust. Soc. Am. 9, 11-23.

Stevens, S.S. (1972). "Perceived level of noise by Mark VII and decibels (E)," J. Acoust. Soc. Am. 51, 575-601.

Studebaker, G.A. (1977). "Utilization of real-time spectral analysis for the electroacoustic evaluation of hearing aids," Presented at the conference on Current Trends in Auditory and Hearing Prosthetics Research, July 25-27, 1977, Univ. of Wyoming (Laramie, Wy.).

Stutz, R. (1975). "A third dimension," Hearing Aid J. 28:3, 19, 36.

Termin, F.E. and Pettit, J.M. (1952). Electronic Measurements (McGraw-Hill, N.Y.) p. 341.

Thann, W.K. and Tracy, K. (1979). "Names of writer's children inserted just for fun," Antipedantic Journal 2, 1311.

Thompson, G. and Lassman, F. (1970). "Listener preference for selective vs. flat amplification for a high-frequency hearing-loss population," J. Speech Hear. Res. 13, 670-672.

Tillman, T.W. (1962). "A critical view of the SAL test," Tech. Docum. Rep. No. SAM-TDR-62-96, School of Aerospace Medicine, Aerospace Medical Div. (AFSC), U.S. Air Force, (Brooks Air Force Base, TX).

Tillman, T.W. and Carhart, R. (1966). "An expanded test for speech discrimination utilizing CNC monosyllabic words (Northwestern University Auditory Test No. 6), technical report SAM-TR-66-55. USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC), (Brooks Air Force Base, TX).

Tillman, T.W., Carhart, R. and Olsen, W.O. (1970). "Hearing aid efficiency in a competing speech situation," J. Speech Hear. Res. 13, 789-811.

Toepholm, C. (1979). "Increasing demands on hearing aids and the ETYMOTIC GAIN," Hearing Aid J. 32:3 (Jan. 1979), 8, 38-39.

Toole, F.E. (1978). "Earphones and ears: getting the act together," Presented at Midwest Acoustics Conference, April 8, 1978, Northwestern Univ. (Evanston, IL.).

Trimmer, J.D. and Firestone, F.A. (1937). "Investigation of Subjective Tones by Means of the Steady Tone Phase Effects," J. Acoust. Soc. Am. 9, 24-29.

Victoreen, J.A. (1960). Hearing Enhancement (Charles Thomas, Springfield).

Victoreen, J.A. (1973). Basic Principles of Otometry (Charles Thomas, Springfield).

Villchur, E. (1962). "A method of testing loudspeakers with random noise input," J. Audio Eng. Soc. 10, 306-309.

Villchur, E. (1964). "Technique of masking live versus recorded comparisons," Audio (October, 1964), 34.

Villchur, E. (1973). "Signal processing to improve speech intelligibility in perceptual deafness," J. Acoust. Soc. Am. 53, 1646-1657.

Villchur, E. (1978). "A critical survey of research on amplitude compression," in Sensorineural hearing impairment and hearing aids (eds. Ludvigsen, C. and Barfod, J.) Scand. Audiol. Suppl 6, 305-314.

Wallach, H. (1940). "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol., 27, 339-368.

Wallenfels, H.G. (1967). Hearing Aids on Prescription (Charles Thomas, Springfield).

Wang, C.-Y. and Dallos, P. (1972). "Latency of whole nerve action potentials: influence of hair cell normalcy," J. Acoust. Soc. Am. 52, 1678-1686.

Watson, L.A. (1944). "Certain fundamental principles in prescribing and fitting hearing aids," Laryngoscope 54, 531-557.

Watson, L.A. and Tolan, T. (1949). Hearing Tests and Hearing Instruments (Williams and Wilkins, Baltimore: reprinted by Hafner, N.Y., 1967).

Watson, N.A. and Gales, R.S. (1943). "Bone-conduction threshold measurements: Effects of occlusion, enclosures, and masking devices," J. Acoust. Soc. Am. 14, 207-215.

Watson, N.A. and Knudsen, V.O. (1940). "Selective Amplification in Hearing Aids," J. Acoust. Soc. Am. 11, 406-419.

Wegel, R.L. and Lane, C.E. (1924). "The Auditory Masking of One Pure Tone by Another and its Probable Relation to the Dynamics of the Inner Ear," Rev. 23, 266-285.

Wenner, C.H. (1968). "Intensities of aural difference tones," J. Acoust. Soc. Am. 43, 77-80.

Wever, E.G. and Lawrence, M. (1954). Physiological Accustics (Princeton Univ. Press, Princeton).

Wiener, F.M. and Ross, D.A. (1946). "The pressure distribution in the auditory canal in a progressive sound field," J. Acoust. Soc. Am. 18, 401-408.

Zwicker, E. (1955). "Der ungewohnliche amplitudengang der nichtlinearen verserrungen des ohres," (The usual amplitude behavior of nonlinear distortion of the ear), Acustica 5, 67-74.

Zwicker, E. (1958). "Der kubische differenzton und die erregung des Gehors," Acustica 20, 206-209.

Zwicker, E. (1976). "Psychoacoustic Equivalent of Period Histograms," J. Acoust. Soc. Amer. 59, 166-175.

Zwislocki, J.J. (1970). "An acoustic coupler for earphone calibration," Report LSC-S-7, Lab. of Sensory Comm., Syracuse Univ. (Syracuse, N.Y.).

APPENDIX A: COMPUTER PROGRAMS·

This appendix contains the listing of the SORT and

MANOV3 computer programs, which were written in the FORTRAN

language, as well as a sample computer session to illustrate

the use of the two programs. Handwritten comments are

included in some places to facilitate understanding or

warn of pitfalls in application.

NOTE

The MANOV3 program was written to maximize
ease of debugging rather than to minimize source
code. The basic defining formulae for the sums
of squares (Brownlee, P 507) were used directly,
for example, rather than the traditional forms
which are computationally more effecient. As a
result, the program is virtually self documenting,
but the source code is at least twice as long as
would be required for an "efficient" FORTRAN
program.

SORT: RANDOMIZED-BLOCK DECODING

```
10    DIMENSION NFILE(4), IBLK(8,12),ISORT(8,12)
100   WRITE(9,110)
110 110 FORMAT('INPUT NO. OF SYSTEMS AND NO. OF BLOCKS: IM (ROWS)'
111 + ' AND JM (COLUMNS)'/)
120 READ(9,255) FIM,FJM
130   IM=FIM
135   JM=FJM
200   DO 290 J=1,JM
210   WRITE(9,220)J
220 220 FORMAT(//,'INPUT DATA CODING SCHEME IN BLOCK NUMBER',I5,
221 + /'COMPARISON/SYSTEM NO.')
230   DO 290 I=1,IM
235   IO=I-1
240   WRITE(9,245)IO
245 245 FORMAT(I4)
250   READ(9,255) D
255 255 FORMAT(2F12.0)
260   IBLK(I,J)=D
290 290 CONTINUE
300   DO 330 J=1,JM
310   DO 330 I=1,IM
320   DO 330 IT=1,IM
330 330 IF(IBLK(IT,J).EQ.I) ISORT(I,J)=IT
345   WRITE(9,350)(JT,JT=1,JM)
350 350 FORMAT('DATA CODING MATRIX FOR SYSTEM NUMBER:'/
351 + 'COMPARISON             BLOCK NUMBER',/
352 + 'NUMBER    ',12(2X,I4),/)
360   DO 390 I=1,IM
370   IO=I-1
374 *
375 * IO FOLLOWS THE ORIG. COMP. NUMBRNG WHERE COMP
376 * #1 IS FIRST
377 *
380   WRITE(9,385)(IO,(IBLK(I,J),J=1,JM))
385 385 FORMAT(I7,3X,12(2X,I4),//)
390 390 CONTINUE
400   WRITE(9,405)
405 405 FORMAT('OUTPUT FILE=')
410   READ(9,415)(NFILE(M),M=1,3)
415 415 FORMAT(4A2)
420   NFILE(4)='  '
430   CALL DEFINE(1,NFILE)
435   DO 440 J=1,JM
440 440 WRITE(1,445)(J,(ISORT(I,J),I=1,IM))
445 445 FORMAT(I7,3X,12(2X,I4))
500   DO 510 J=1,JM
510 510 WRITE(9,445)(J,(ISORT(I,J),I=1,IM))
780   WRITE(9,790)(NFILE(M),M=1,3)
790 790 FORMAT('DONT FORGET TO SAVE ISORT FILE = ',4A2)
799   STOP
800   END
```

MANOV3: 3-WAY ANALYSIS OF VARIANCE

```
1 *MANOV3 78/10/9    MEAD KILLION
2 * USE 'FORTRAN MANOV3 L X' TO OBTAIN ARCSIN TRANSFORMATION OF DATA
5    INTEGER ALPHA,ALPHB
6    REAL INTXO
9    COMMON X(8,12,30)
10     DIMENSION NFILE(4),DX(8,12),ISORT(8,12)
11   DIMENSION LREP(12,12),REP(12),SSQER(12),LM(12),ERRK(30)
12   DIMENSION INTXO(13),INTX(13),WC(8,12)
14   DIMENSION XI(8),XJ(12),XK(30),XIJ(8,12),XIK(8,30),XJK(12,30)
15   DIMENSION ALPHA(8),ALPHB(8),BETA(8),GAM(8)
16   DATA ((ALPHA(IA),ALPHB(IA)),IA=1,8)/2H  ,2H A,2H  ,2H B,
17 + 2H  ,2H C,2H  ,2HAB,2H  ,2HAC,2H  ,2HBC,2H A,2HBC,2HER,2HR./
18   WRITE (9,31)
19   OPT=0
20 20 DO 30 I=1,8
21   DO 30 J=1,12
22   DX(I,J)=0.
23   XIJ(I,J)=0.
24   WC(I,J)=0.
29 29 DO 30 K=1,30
30 30 X(I,J,K)=0.
31 31 FORMAT(' MANOV3  VERSION 78/10/9;    MEAD    ',/)
32   DO 35 J=1,11
33   LM(J)=0
34   DO 35 L=1,11
35 35 LREP(J,L)=0
37   CALL OPREFX(0)
38   IF(OPT.EQ.1.)GO TO 70
40   WRITE(9,41)
41 41 FORMAT('DATA INPUT FILE=')
44   READ (9,45)(NFILE(M),M=1,3)
45 45 FORMAT(4A2)
46   NFILE(4)=' '
48   CALL DEFINE(2,NFILE)
50   WRITE (9,51)
51 51 FORMAT('DATA CODING FILE=')
54   READ (9,45)(NFILE(M),M=1,3)
70 70 LMX=1
71   JNMX=0
72   DO 86 J=1,11       .
73   WRITE(9,74)J
74 74 FORMAT('DATA COLUMNS AVERAGED IN COLUMN',I4,' ARE')
76   READ (9,77)(REP(L),L=1,6)
77 77 FORMAT (6F12.0)
78   IF (REP(1).EQ.0.) GO TO 88
80   JNMX=JNMX+1
82   DO 86 L=1,6
83   IF(REP(L).NE.0.)LM(J)=L
84   IF(LM(J).GT.LMX)LMX=LM(J)
86 86 LREP(J,L)=REP(L)
88 88 CONTINUE
100  100 ICR=:000215
102   CALL OPREFX(ICR)
120   READ(2,125)FLN,FIM,FJM,FKM,FLCK
125 125 FORMAT(5F12.0)
130   IM=FIM
```

PAGE   2

MANOV3: 3-WAY ANALYSIS OF VARIANCE

```
132   JM=FJM
134   KM=FKM
140   DO 144 J=1,JM
141   DO 144 I=1,IM
144 144 ISORT(I,J)=I
148   IF ((NFILE(1).EQ.'NO').OR.(NFILE(1).EQ.'  '))GO TO 165
150   CALL DEFINE(1,NFILE)
159   DO 165 J=1,JM
160   READ(1,161)M,(ISORT(I,J),I=1,IM)
161 161 FORMAT(I7,3X,10(2X,I4))
162   IF(M.NE.J) WRITE(9,163)J,M
163 163 FORMAT('COLUMN #',I4,'FINDS LINE #',I4)
165 165 CONTINUE
168 DO 210 K=1,KM
170   READ(2,125)FLN,FK
171   KIN=FK
173   IF((KIN.NE.K).AND.(FLCK.EQ.1.))WRITE(9,175)FLN,FK,K
175 175 FORMAT('LINE NO.',F9.0,';',F9.0,'=SUBJECT NO. ',I4,'?')
180   DO 190 J=1,JM
190 190 READ(2,195)FJNK,(DX(I,J),I=1,IM)
195 195 FORMAT(8F12.0)
202   DO 210 J=1,JM
204   DO 210 I=1,IM
206   IST=ISORT(I,J)
208 . A=SQRT(DX(IST,J)/100.)
209 . DX(IST,J)=100./1.5708*ATAN(A/SQRT(1.000001-A*A))
210 210 X(I,J,K) = DX(IST,J)
214   IF(JNMX.EQ.0) GO TO 268
216   ERR=0.
218   DO 266 K=1,KM
220   DO 230 J=1,JNMX
222   DO 230 I=1,IM
224   DX(I,J)=0.
225   LMNOW=LM(J)
226   DO 228 L=1,LMNOW
227   JNOW=LREP(J,L)
228 228 DX(I,J)=DX(I,J) + X(I,JNOW,K)
230 230 DX(I,J) = DX(I,J)/LM(J)
236   JDV=0
238   ERRK(K)=0.
240   DO 258 J=1,JNMX
242   IF(LM(J).EQ.1) GO TO 258
243   JDV=JDV+1
244   SSQ=0.
246   DO 254 I=1,IM
248   SS=0.
249   LMNOW=LM(J)
250   DO 252 L=1,LMNOW
251   JNOW=LREP(J,L)
252 252 SS=SS+(X(I,JNOW,K)-DX(I,J))**2
254 254 SSQ=SSQ+SS
256 256 ERRK(K)=ERRK(K)+SSQ/(LM(J)-1)
258 258 CONTINUE
259   IF (JDV.NE.0) ERRK(K)=ERRK(K)/(JDV*IM)
260 260 ERR=ERR+ERRK(K)*IM*JDV*(LMX-1)
261 * NOTE THAT ERR AND ERRMS ARE CALCULATED ONLY FOR REPLICATED CLMNS
```

PAGE   3

MANOV3:   3-WAY ANALYSIS OF VARIANCE

```
262    IF (LMX.EQ.1)GO TO 264
263    ERRMS=ERR/(IM*JDV*KM*(LMX-1))
264 264 DO 266 I=1,IM
265    DO 266 J=1,JNMX
266 266 X(I,J,K)=DX(I,J)
267    JM=JNMX
268 268 CONTINUE
270 * LINES 270-299 RESERVED FOR MATH. TRANSF. OF DATA
302    DO 309 I=1,IM
303    XI(I)=0.
304    DO 308 J=1,JM
306    DO 308 K=1,KM
308 308 XI(I)=XI(I)+X(I,J,K)
309 309 XI(I)=XI(I)/(JM*KM)
312    DO 319 J=1,JM
313    XJ(J)=0.
314    DO 318 I=1,IM
316    DO 318 K=1,KM
318 318 XJ(J)=XJ(J)+X(I,J,K)
319 319 XJ(J)=XJ(J)/(IM*KM)
322    DO 329 K=1,KM
323    XK(K)=0.
324    DO 328 I=1,IM
326    DO 328 J=1,JM
328 328 XK(K)=XK(K)+X(I,J,K)
329 329 XK(K)=XK(K)/(IM*JM)
332    DO 339 I=1,IM
333    DO 339 J=1,JM
334    XIJ(I,J)=0.
336    DO 338 K=1,KM
338 338 XIJ(I,J)=XIJ(I,J)+X(I,J,K)
339 339 XIJ(I,J)=XIJ(I,J)/KM
342    DO 349 J=1,JM
343    DO 349 K=1,KM
344    XJK(J,K)=0.
345    DO 348 I=1,IM
348 348 XJK(J,K)=XJK(J,K)+X(I,J,K)
349 349 XJK(J,K)=XJK(J,K)/IM
352    DO 359 I=1,IM
353    DO 359 K=1,KM
354    XIK(I,K)=0.
356    DO 358 J=1,JM
358 359 XIK(I,K)=XIK(I,K)+X(I,J,K)
359 359 XIK(I,K)=XIK(I,K)/JM
361    XBAR=0.
362    DO 368 I=1,IM
363    DO 368 J=1,JM
364    DO 368 K=1,KM
368 368 XBAR=XBAR+X(I,J,K)
369    XBAR=XBAR/(IM*JM*KM)
502    A=0.
504    DO 507 I=1,IM
507 507 A=A+(XI(I)-XBAR)**2
508    A=JM*KM*A*LMX
509    AMS=A/(IM-1)
512    B=0.
```

PAGE   4

MANOV3:   3-WAY ANALYSIS OF VARIANCE

```
513    IF(JM.EQ.1) GO TO 520
514    DO 517 J=1,JM
517 517 B=B+(XJ(J)-XBAR)**2
518    B=IM*KM*LMX*B
519    BMS=B/(JM-1)
520 520    IF(KM.EQ.1)GO TO 532
522    C=0.
524    DO 527 K=1,KM
527 527 C=C+(XK(K)-XBAR)**2
528    C=IM*JM*LMX*C
529    CMS=C/(KM-1)
532 532 AB=0.
533    IF (JM.EQ.1)GO TO 540
534    DO 537 I=1,IM
535    DO 537 J=1,JM
537 537 AB=AB+(XIJ(I,J)-XI(I)-XJ(J)+XBAR)**2
538    AB=KM*LMX*AB
539    ABMS=AB/((IM-1)*(JM-1))
540 540 IF(KM.EQ.1)GOTO 600
542    AC=0.
544    DO 547 I=1,IM
545    DO 547 K=1,KM
547 547 AC=AC+(XIK(I,K)-XI(I)-XK(K)+XBAR)**2
548    AC=JM*LMX*AC
549    ACMS=AC/((IM-1)*(KM-1))
551    IF(JM.EQ.1)GO TO 570
552    BC=0.
554    DO 557 J=1,JM
555    DO 557 K=1,KM
557 557 BC=BC+(XJK(J,K)-XJ(J)-XK(K)+XBAR)**2
558    BC=IM*LMX*BC
559    BCMS=BC/((JM-1)*(KM-1))
562    ABC=0.
563    DO 566 I=1,IM
564    DO 566 J=1,JM
565    DO 566 K=1,KM
566 566 ABC=ABC+(X(I,J,K)-XIJ(I,J)-XJK(J,K)-XIK(I,K)+XI(I)+XJ(J)
567 +  +XK(K)-XBAR)**2
568    ABC=LMX*ABC
569    ABCMS=ABC/((IM-1)*(JM-1)*(KM-1))
570 570 WCEL=0.
572    DO 579 I=1,IM
573    DO 579 J=1,JM
574    WC(I,J)=0.
576    DO 577 K=1,KM
577 577 WC(I,J)=WC(I,J)+(X(I,J,K)-XIJ(I,J))**2
578    WCEL=WCEL+WC(I,J)
579 579 WC(I,J)=SQRT(WC(I,J)/(KM-1.))
580    WCMS=WCEL/(IM*JM*(KM-1))
582    WCSTDE=SQRT(WCMS)
600 600 CONTINUE
810    WRITE(9,820)
820 820 FORMAT(//'SYSTEM',20X,'BLOCK NUMBER',29X,'SYSTEM',/
821 +  'NUMBER',3X,'1     2     3     4     5     6     7     8     9    10'
822 +  '    11    12    AVG.')
830    DO 840 I=1,IM
```

MANOV3: 3-WAY ANALYSIS OF VARIANCE

```
832    DO 834 J=1,12
834 834 INTXO(J)=XIJ(I,J)
836    INTXO(13)=XI(I)
840 840 WRITE(9,845)I,(INTXO(J),J=1,13)
845 845 FORMAT(/,I4,2X,12F5.1,2X,F5.1)
850    DO 852 J=1,JM
852 852 INTXO(J)=XJ(J)
854    INTXO(13)=XBAR
860    WRITE(9,865)(INTXO(J),J=1,13)
865 865 FORMAT('BLOCK',/,' AVGS ',12F5.1,2X,F5.1)
869    IF(KM.EQ.1)GO TO 895
870    WRITE(9,875)
875 875 FORMAT(///'STD.DEV. IN EACH CELL ')
882    DO 890 I=1,IM
883    DO 884 J=1,JM
884 884 INTXO(J)=WC(I,J)
890 890 WRITE(9,845)I,(INTXO(JTM),JTM=1,JM)
892    WRITE(9,893)WCMS,WCSTDE
893 893 FORMAT('W/CELLS MEAN SQ = ',F15.2,'  STD.DEV.',F10.2)
895 895 CONTINUE
900    BETA(1)=A
901    BETA(2)=B
903    BETA(3)=C
904    BETA(4)=AB
905    BETA(5)=AC
906    BETA(6)=BC
907    BETA(7)=ABC
908    BETA(8)=ERR
911    GAM(1)=AMS
912    GAM(2)=BMS
913    GAM(3)=CMS
914    GAM(4)=ABMS
915    GAM(5)=ACMS
916    GAM(6)=BCMS
917    GAM(7)=ABCMS
918    GAM(8)=ERRMS
920    INTX(1)=IM-1
921    INTX(2)=JM-1
923    INTX(3)=KM-1
924    INTX(4)=(IM-1)*(JM-1)
925    INTX(5)=(IM-1)*(KM-1)
926    INTX(6)=(JM-1)*(KM-1)
927    INTX(7)=(IM-1)*(JM-1)*(KM-1)
928    INTX(8)=IM*JDV*KM*(LMX-1)
930    WRITE(9,945)
945 945 FORMAT(///'   SOURCE OF VARIATION',3X,'SUM OF SQUARES',3X,
946 + 'DEGREES OF FREEDOM',3X,'MEAN SQUARE'/)
950    DO 960 IA=1,8
951    IF((JM.EQ.1).AND.(IA.EQ.2))IA=3
952    IF((KM.EQ.1).AND.(IA.EQ.3))IA=4
953    IF((JM.EQ.1).AND.(IA.EQ.4))IA=5
954    IF((JM.EQ.1).AND.(IA.EQ.6))IA=8
955    IF((KM.EQ.1).AND.(IA.GT.4).AND.(IA.LT.8))GO TO 960
956 956 IF((LMX.EQ.1).AND.(IA.EQ.8))GO TO 970
959 959 WRITE(9,965) ALPHA(IA),ALPHB(IA),BETA(IA),INTX(IA),GAM(IA)
960 960 CONTINUE
```

MANOV3: 3-WAY ANALYSIS OF VARIANCE

```
965 965 FORMAT(15X,2A2,4X,F15.2,6X,I6,7X,F15.2)
970 970 CONTINUE
972    IF(LMX.GT.1)WRITE(9,973)LMX
973 973 FORMAT('NOTE:   THE ANAL. OF VAR. ASSUMES ALL COLUMNS CONTAIN '
974 + 'THE SAME'/' NUMBER (',I3,') OF REPLICATES.  IF NOT, ONLY THE '
975 + 'ERR. SUM-OF-SQUARES'/' AND MEAN-SQUARE VALUES ARE STRICTLY '
976 + 'CORRECT.')
980 .    WRITE(9,981)
981 . 981 FORMAT('NOTE: ALL DATA ARCSIN TRANSFORMED ')
1100 1100 WRITE(9,1102)
1102 1102 FORMAT('OPTION')
1104    READ(9,1106) OPT
1106 1106 FORMAT(F12.0)
1108    IF(OPT.NE.2)GOTO 1140
1110 1110 WRITE(9,1111)
1111 1111 FORMAT('INDIVIDUAL MEAN SQUARES BY SUBJECT NUMBER')
1114    DO 1120 K=1,KM
1120 1120 WRITE(9,1121)K,ERRK(K)
1121 1121 FORMAT(I4,F15.2)
1130    GO TO 1100
1140 1140 REWIND 2
1141    IF((NFILE(1).NE.'NO').AND.(NFILE(1).NE.'  '))REWIND 1
1140    CALL OPREFX(0)
1150    WRITE(9,31)
1110    GO TO 20
2001 END
```

ON THIS AND THE SUCCEEDING PAGES, A SAMPLE COMPUTER SESSION
(RUN ON A HONEYWELL 1648 TIME-SHARING SYSTEM) IS SHOWN TO ILLUSTRATE
THE USE OF THE SORT AND MANOV3 PROGRAMS. FOR THIS EXAMPLE, THE
AVERAGE DATA OBTAINED FROM EACH OF THE THREE SUBJECT GROUPS FOR
THE THREE PROGRAM-SELECTION BLOCKS ON COMPARISON TAPE A WERE
USED (SEE APPENDIX D), WITH EACH SUBJECT GROUP TREATED AS A SINGLE
SUBJECT.

?FORTRAN SORT CSORT   ◄── COMPILE   (Underlined text is Operator Entry)
?LOAD (OBSORT) CSORT   ◄── LOAD (AND RUN)

INPUT NO. OF SYSTEMS AND NO. OF BLOCKS: IM (ROWS) AND JM (COLUMNS)
! 7,3

INPUT DATA CODING SCHEME IN BLOCK NUMBER   1
COMPARISON/SYSTEM NO.
   0! 1

   1! 2

   2! 3

   3! 4

   4! 5

   5! 6

   6! 7

INPUT DATA CODING SCHEME IN BLOCK NUMBER   2
COMPARISON/SYSTEM NO.
   0! 1

   1! 5

   2! 6

   3! 3

   4! 7

   5! 4

   6! 2

INPUT DATA CODING SCHEME IN BLOCK NUMBER   3
COMPARISON/SYSTEM NO.
   0! 1

   1! 4

   2! 2

   3! 7

   4! 3

   5! 5

   6! 6

DATA CODING MATRIX FOR SYSTEM NUMBER;
COMPARISON   BLOCK NUMBER
NUMBER   1   2   3
   0   1   1   1
   1   2   5   4
   2   3   6   2
   3   4   3   7
   4   5   7   3
   5   6   4   5
   6   7   2   6
OUTPUT FILE=! A

   1   1   2   3   4   5   6   7
   2   1   7   4   6   2   3   5
   3   1   3   5   2   6   7   4
DONT FORGET TO SAVE ISORT FILE = A
STOP,
? CREATE EXMPL

BEGIN
Line
Numbers ┌── Number of Rows (Systems)
 │   ┌── Number of Columns (Program-selection Blocks)
 │ │ ┌── Number of Subjects
1,7,3,5
10,1 ◄── Subject Number
11,25,65,81,72,57,73,74 ⎤ Data From "Subject" (Untrained Listeners
12,14,622,,86,83,88,80,57 ⎬ Averages) Number 1, Comparisons on Tape A
13,24,76,58,65,88,69,93 ⎦
20,2
21,14,59,83,74,52,66,54
22,9,50,77,86,66,75,49
23,22,80,52,66,96,50,92
30,3
31,18,59,93,79,56,78,68
32,12,59,87,96,74,68,48
33,20,84,48,75,99,72,98 CRCR ◄── Two Carriage Returns denote end
                              of File. (A single CR — not shown —
                              denotes end of data line).

NOTE: OPERATOR ENTRIES
OF DATA NOT
UNDERLINED HERE TO
AVOID CLUTTER

?LRUN M

   MANOV3   VERSION 78/10/9;   MEAD
DATA INPUT FILE=! EXMPL
DATA CODING FILE=! A
DATA COLUMNS AVERAGED IN COLUMN   1 ARE!

STOP, READ EOF FILE 2   ◄── Common cause of abort: Number of
?PATCH EXMPL   Data Sets differed from number of Ss.

BEGIN   ┌── Only 3 "subjects" instead of 5 typed in above
1,7,3,3

? LIST EXMPL

1,7,3,3
10,1
11,25,65,81,72,57,73,74
12,14,62,86,83,88,80,57
13,24,76,58,65,88,69,93
20,2
21,14,59,83,74,52,66,54
22,9,50,77,86,66,75,49
23,22,80,52,66,96,50,92
30,3
31,18,59,93,79,56,78,68
32,12,59,87,96,74,68,48
33,20,84,48,75,99,72,98

?LRUN M `NAME GIVEN (PREVIOUSLY COMPILED AND LOADED) OBJECT -FILE`
`VERSION OF "MANOV3"`
MANOV3 VERSION 78/10/9; HEAD
DATA INPUT FILE=! EXMPL `FORTRAN SOURCE FILE.`
DATA CODING FILE=! A
DATA COLUMNS AVERAGED IN COLUMN 1 ARE!

SYSTEM NUMBER	BLOCK NUMBER 1	2	3	4	5	6	7	8	9	10	11	12	SYSTEM AVG.
1	19.0	11.7	22.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	17.6
2	61.0	51.3	52.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	55.0
3	85.7	88.3	94.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	89.4
4	75.0	74.3	80.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	76.4
5	55.0	57.0	63.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	58.6
6	72.3	83.3	94.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	83.3
7	65.3	76.0	68.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	70.0
BLOCK AVGS	61.9	63.1	68.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	64.3

STD.DEV. IN EACH CELL

	1	2	3
1	5.6	2.5	2.0
2	3.5	4.9	5.0
3	6.4	6.8	5.7
4	3.6	6.0	4.0
5	2.6	6.2	11.9
6	6.0	5.5	3.2
7	10.3	11.1	5.5

W/CELLS MEAN SQ = 38.73 STD.DEV. 6.22

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE
A	31311.11	6	5218.52
B	428.67	2	214.33
C	445.81	2	222.91
AB	1109.55	12	92.46
AC	669.08	12	55.76
BC	111.52	4	27.88
ABC	400.25	24	16.68

OPTION!

APPENDIX B: THREE-WAY ANALYSIS OF VARIANCE OF UNTRAINED-LISTENER DATA

This appendix contains the computer printout of a three-way analysis of variance applied to the Untrained-Listener data. The first printout is for the raw rating data, while the second is for the data after application of the arcsin transform. The arcsin transform has been normalized to a zero-to-100 percent scale by multiplying the transformed data by 200/pi. (The exact transform is given in lines 208 and 209 of the MANOV3 program listing in Appendix A. The form used here is equivalent to the conventional arcsin transform as given, for example, in Brownlee [1965, P145]).

Descriptive overlays have been added to facilitate the examination of the data.

MANOV3 VERSION 78/10/9;   MEAD
DATA INPUT FILE=! KEBOTH *(Untrained Listeners)*   ┌── *First Day Ratings*
DATA CODING FILE=! CAND2   ┌── *Second Day Ratings*
DATA COLUMNS AVERAGED IN COLUMN   1 ARE! 1,7 (ORCHESTRA, FORTISSIMO)
DATA COLUMNS AVERAGED IN COLUMN   2 ARE! 2,8 (PIANO TRIO, CHORUS #6)
DATA COLUMNS AVERAGED IN COLUMN   3 ARE! 3,9 (WIDEBAND NOISE)
DATA COLUMNS AVERAGED IN COLUMN   4 ARE! 4,10 (PIANO TRIO, CHORUS #7)
DATA COLUMNS AVERAGED IN COLUMN   5 ARE! 5,11 (MALE VOICE)
DATA COLUMNS AVERAGED IN COLUMN   6 ARE! 6,12 (ORCHESTRA, FORTE)
DATA COLUMNS AVERAGED IN COLUMN   7 ARE!

| SYSTEM | | | | BLOCK NUMBER | | | | | | | | | SYSTEM |
NUMBER	1	2	3	4	5	6	7	8	9	10	11	12	AVG.
1	8.1	7.8	12.9	12.5	23.4	11.0		POCKET RADIO					12.6
2	54.3	30.7	34.4	39.6	57.8	65.5		DISCOUNT STEREO					47.1
3	56.8	31.4	41.3	37.5	51.5	76.8		SPEECH AUDIOMETER					49.2
4	66.8	63.2	43.0	60.9	54.6	67.4		POPULAR HEADPHONES					59.3
5	81.4	76.7	53.3	69.0	80.8	75.1		MONITOR SPEAKERS					72.7
6	80.6	67.1	65.6	71.0	86.4	83.6		ITE HEARING AIDS					75.7
7	87.1	59.6	73.1	61.8	87.5	82.9		OTE HEARING AIDS					75.4
BLOCK AVGS	62.2	48.1	46.2	50.3	63.1	66.1	0.0	0.0	0.0	0.0	0.0	0.0	56.0

STD.DEV. IN EACH CELL

1	9.5	7.8	15.0	13.9	19.7	12.9
2	18.2	14.5	21.3	16.9	18.6	16.1
3	16.6	15.4	15.7	18.4	23.0	15.8
4	12.8	21.6	26.1	16.4	27.8	19.4
5	12.3	15.0	20.2	10.6	12.9	15.5
6	9.9	19.5	16.1	15.4	11.0	12.7
7	7.8	19.7	11.9	16.8	11.6	11.6

W/CELLS MEAN SQ =   269.78 STD.DEV.   16.43

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE
A	881628.50	6	146938.09
B	127824.58	5	25564.91
C	125368.20	23	5450.79
AB	95620.20	30	3187.34
AC	99674.34	138	722.28
BC	94577.92	115	822.42
ABC	201600.94	690	292.18
ERR.	245284.50	1008	243.34

LRUN MAS ◄── *Name given Arcsin Version*

MANOV3 VERSION 78/10/9;   MEAD   ┌── *First Day Ratings*
DATA INPUT FILE=! KEBOTH *(Untrained Listeners)*   ┌── *Second Day Ratings*
DATA CODING FILE=! CAND2
DATA COLUMNS AVERAGED IN COLUMN   1 ARE! 1,7 (ORCHESTRA, FORTISSIMO)
DATA COLUMNS AVERAGED IN COLUMN   2 ARE! 2,8 (PIANO TRIO, CHORUS #6)
DATA COLUMNS AVERAGED IN COLUMN   3 ARE! 3,9 (WIDEBAND NOISE)
DATA COLUMNS AVERAGED IN COLUMN   4 ARE! 4,10 (PIANO TRIO, CHORUS #7)
DATA COLUMNS AVERAGED IN COLUMN   5 ARE! 5,11 (MALE VOICE)
DATA COLUMNS AVERAGED IN COLUMN   6 ARE! 6,12 (ORCHESTRA, FORTE)
DATA COLUMNS AVERAGED IN COLUMN   7 ARE!

| SYSTEM | | | | BLOCK NUMBER | | | | | | | | | SYSTEM |
NUMBER	1	2	3	4	5	6	7	8	9	10	11	12	AVG.
1	13.3	13.4	18.7	18.3	28.7	15.8		POCKET RADIO					18.0
2	53.5	35.8	36.3	42.0	55.4	61.5		DISCOUNT STEREO					47.4
3	54.8	36.6	43.9	40.8	51.2	71.9		SPEECH AUDIOMETER					49.9
4	62.0	60.2	44.1	57.8	53.1	63.3		POPULAR HEADPHONES					56.8
5	73.8	70.6	52.2	63.0	74.9	69.9		MONITOR SPEAKERS					67.4
6	73.5	63.9	60.6	66.0	79.6	77.4		ITE HEARING AIDS					70.2
7	79.6	55.7	66.1	59.4	82.0	76.1		OTE HEARING AIDS					70.0
BLOCK AVGS	58.6	48.2	46.0	49.6	60.7	62.3	0.0	0.0	0.0	0.0	0.0	0.0	54.2

STD.DEV. IN EACH CELL

1	12.0	10.6	14.5	13.8	18.4	13.8
2	12.9	11.2	19.4	13.0	13.0	13.2
3	11.6	11.8	11.0	13.7	16.5	14.1
4	9.2	16.9	20.4	11.6	22.1	14.7
5	11.0	12.7	13.8	7.4	13.0	13.3
6	9.1	16.5	12.3	13.1	10.5	13.4
7	9.9	14.0	8.5	14.2	13.0	11.1

W/CELLS MEAN SQ =   184.46 STD.DEV.   13.58

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE
A	592465.37	6	9??4.23
B	84371.98	5	16?74.40
C	81576.86	23	3546.82
AB	62036.37	30	2067.88
AC	72521.33	138	525.52
BC	66575.91	115	578.92
ABC	135707.47	690	196.68
ERR.	172427.12	1008	171.06

NOTE: ALL DATA ARCSIN TRANSFORMED
OPTION!

APPENDIX C: THREE-WAY ANALYSIS OF VARIANCE OF

GOLDEN-EAR AND TRAINED-LISTENER DATA

This appendix contains the computer printout of a three-way

analysis of variance applied to the ratings obtained from the

Golden-Ear subjects (DATA INPUT FILE = GLDN) and the Trained-

Listener subjects (DATA INPUT FILE = TECH). The analysis was

performed on the raw (untransformed) data.

```
MANOV3  VERSION 78/10/9;   MEAD
DATA INPUT FILE=! GLDN
DATA CODING FILE=! CAND
DATA COLUMNS AVERAGED IN COLUMN   1 ARE!
```

SYSTEM NUMBER	1	2	3	4	BLOCK NUMBER 5	6	7	8	9	10	11	12	SYSTEM AVG.
1	13.0	16.0	15.0	22.0	27.0	13.0		POCKET RADIO					17.7
2	50.0	34.0	29.0	39.0	52.0	51.0		DISCOUNT STEREO					42.5
3	46.0	38.0	33.0	38.0	42.0	56.0		SPEECH AUDIOMETER					42.2
4	46.0	44.0	29.0	42.0	41.0	59.0		POPULAR HEADPHONES					43.5
5	63.0	71.0	56.0	56.0	86.0	73.0		MONITOR SPEAKERS					67.5
6	71.0	71.0	63.0	69.4	90.4	83.0		ITE HEARING AIDS					74.6
7	73.0	67.6	67.0	71.0	93.2	83.6		OTE HEARING AIDS					75.9
BLOCK AVGS	51.7	48.8	41.7	48.2	61.7	59.8	0.0	0.0	0.0	0.0	0.0	0.0	52.0

STD.DEV. IN EACH CELL

									BLOCK CODE
1	5.7	6.5	8.7	13.0	13.0	6.7			1. ORCHESTRA, FORTISSIMO
2	20.3	13.9	12.4	10.8	9.1	13.4			2. PIANO TRIO, CHORUS #6
3	9.6	17.9	12.0	13.5	17.9	16.4			3. WIDEBAND NOISE
4	20.1	19.7	12.4	14.4	14.3	11.4			4. PIANO TRIO, CHORUS #7
5	16.4	15.2	10.8	16.4	11.9	14.8			5. MALE VOICE
6	21.0	28.6	11.5	20.2	10.2	14.0			6. ORCHESTRA, FORTE
7	21.1	28.1	17.2	17.8	7.5	19.4			

W/CELLS MEAN SQ =   240.46 STD.DEV.   15.51

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE
A	82850.92	6	13808.49
B	9962.95	5	1992.59
C	21799.88	4	5449.97
AB	5794.45	30	193.15
AC	4499.79	24	187.49
BC	5326.24	20	266.31
ABC	8771.70	120	73.10

SYSTEM NUMBER	1	2	BLOCK NUMBER 3	4	5	6	7	8	9	10	11	12	SYSTEM AVG.
1	10.5	12.0	20.8	14.5	17.8	11.3		POCKET RADIO					14.5
2	59.2	51.5	38.3	55.0	69.2	66.7		DISCOUNT STEREO					56.6
3	68.8	39.2	43.3	47.5	53.3	75.0		SPEECH AUDIOMETER					54.5
4	59.2	59.2	49.2	53.3	55.8	68.3		POPULAR HEADPHONES					57.5
5	77.5	86.7	55.0	86.7	93.3	80.0		MONITOR SPEAKERS					79.9
6	85.8	87.3	68.7	85.8	96.7	92.8		ITE HEARING AIDS					86.2
7	79.2	68.3	65.0	71.7	95.2	88.3		OTE HEARING AIDS					77.9
BLOCK AVGS	62.9	57.7	48.6	59.2	68.8	68.9	0.0	0.0	0.0	0.0	0.0	0.0	61.0

STD.DEV. IN EACH CELL

	1	2	3	4	5	6		BLOCK CODE	
1	6.3	8.2	12.0	8.6	12.0	7.0		1.	ORCHESTRA, FORTISSIMO
2	15.0	7.3	9.8	11.8	13.9	12.1		2.	PIANO TRIO, CHORUS #6
3	11.7	9.2	24.2	13.3	15.1	8.9		3.	WIDEBAND NOISE
4	8.0	4.9	18.6	8.8	8.0	14.0		4.	PIANO TRIO, CHORUS #7
5	5.2	7.5	8.4	6.1	7.5	9.5		5.	MALE VOICE
6	6.6	8.5	13.7	7.4	6.1	6.5		6.	ORCHESTRA, FORTE
7	8.6	8.2	10.5	5.2	5.2	8.2			

W/CELLS MEAN SQ =   109.03 STD.DEV.   10.44

SOURCE OF VARIATION	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE
A	126468.73	6	21078.12
B	12337.95	5	2467.59
C	2692.14	5	538.43
AB	11815.83	30	393.86
AC	2689.30	30	89.64
BC	6074.04	25	242.96
ABC	11439.84	150	76.27

APPENDIX D: COMPLETE LISTING OF ALL COMPARISONS

ON TAPES A, B, C, AND D

This appendix contains a complete listing of all comparisons used in the present experiments, along with the average rating given the system in each comparison by the three subject groups. The listings are organized in sequential order as they appeared on the four comparison tapes A, B, C, and D. The comparisons are grouped by Program-selection blocks. The average fidelity rating from the Untrained-Listener subjects (UL), Golden-Ear subjects (GE), and Trained-Listener subjects (TL) are given for each comparison under the corresponding heading.

DECODING SHEET FOR COMPARISON TAPE A

NOTE: The three blocks on tape A contain different randomizations of the same 7 sound systems.

COMPARISON NUMBER	FIDELITY RATING UL GE TL			COMMENTS (OPTIONAL)
10	25	14	18	Pocket Radio
11	65	59	59	Airline Stereo
12	81	83	93	+3 dB Level Shift
13	72	74	79	Bookshelf Speakers
14	57	52	56	OTE-40 (Hearing Aids with Undamped Tubing)
15	73	66	78	AR3a's in Corners (Difference in sound caused by moving Reference Speakers)
16	74	54	68	Highly Rated Headphones

Oscar Peterson Trio "The Smudge" -- 6th Chorus

COMPARISON NUMBER	FIDELITY RATING UL GE TL			COMMENTS (OPTIONAL)
20	14	9	12	Pocket Radio
21	62	50	59	OTE-40
22	86	77	87	AR3's in Corners
23	83	86	96	+3 dB Level Shift
24	88	66	74	Highly Rated Headphones
25	80	75	68	Bookshelf Speakers
26	57	49	48	Airline Stereo

N.Y. Philharmonic Beethoven Violin Concerto in D ff Passage

COMPARISON NUMBER	FIDELITY RATING UL GE TL			COMMENTS (OPTIONAL)
30	24	22	20	Pocket Radio
31	76	80	84	Bookshelf Speakers
32	58	52	48	Airline Stereo
33	65	66	75	Highly Rated Headphones
34	88	96	99	+3 dB Level Shift
35	69	50	72	OTE-40
36	93	92	96	AR3's in Corners

Male Voice "Joe took father's . . ."

DECODING SHEET FOR COMPARISON TAPE B

NOTES: 1) The first block on tape B contains true "Live vs Recorded" comparisons.
2) The second and third blocks contain the same 7 sound systems.

COMPARISON NUMBER	FIDELITY RATING UL GE TL			COMMENTS (OPTIONAL)
40	27	29	21	Pocket Radio
41	65	77	82	Microphone on KEMAR head (over the ear)
42	72	79	89	Left AR3a vs Live Talker
43	59	56	68	Live Talker: "By speaker" vs "In corner"
44	59	44	56	Left Monitor Speaker vs Live Talker
45	84	80	93	SM-81 Condenser Mics (ORTF position) vs KEMAR (Equal.)
46	77	68	86	SM-81 -3 dB level shift

(Live) Male Voice "Joe took father's . . ."

COMPARISON NUMBER	FIDELITY RATING UL GE TL			COMMENTS (OPTIONAL)
50	7	14	9	Pocket Radio
51	79	90	94	2:1 compression vs none (OTE hearing aid)
52	87	93	98	Two different microphones (OTE hearing aid)
53	71	67	76	Earphone-Earmolds (only) from OTE aid
54	43	21	30	Monophonic (KEMAR left ear only)
55	86	93	92	ITC hearing aids (16 kHz BW)
56	84	83	85	Larger ears on KEMAR (vs standard)

N.Y. Philharmonic Beethoven Violin Concerto in D ff Passage

COMPARISON NUMBER	FIDELITY RATING UL GE TL			COMMENTS (OPTIONAL)
60	7	21	13	Pocket Radio
61	62	63	70	Earphone-Earmold
62	44	19	38	Monophonic
63	74	84	88	ITC hearing aids
64	68	80	83	Larger ears
65	65	80	83	2:1 compression
66	88	92	95	Two different microphones

Oscar Peterson Trio "The Smudge" -- 6th Chorus

DECODING SHEET FOR COMPARISON TAPE C

NOTE: The same 7 systems appear, in randomized order, in all six blocks contained in tapes C and D.

N.Y. Philharmonic Beethoven Violin Concerto in D ff Passage

COMPARISON NUMBER	FIDELITY RATING UL GE TL			COMMENTS (OPTIONAL)
70	8	13	11	Pocket Radio
71	54	50	59	Discount Stereo
72	81	63	78	Monitor Speakers
73	67	46	59	Popular Headphones
74	81	71	86	ITE Hearing Aids (16 kHz BW)
75	58	46	59	Speech Audiometer (TDH-39 earphones)
76	87	73	79	OTE Hearing Aids (8 kHz BW)

Oscar Peterson Trio "The Smudge" -- 6th Chorus

80	8	16	12	Pocket Radio
81	67	71	87	ITE Hearing Aids
82	31	38	39	Speech Audiometer
83	31	34	52	Discount Stereo
84	60	68	68	OTE Hearing Aids
85	63	44	59	Popular Headphones
86	77	71	87	Monitor Speakers

Speech Spectrum Noise 84 dB SPL

90	13	15	21	Pocket Radio
91	42	33	43	Speech Audiometer
92	65	63	69	ITE Hearing Aids
93	54	56	55	Monitor Speakers
94	35	29	38	Discount Stereo
95	73	67	65	OTE Hearing Aids
96	42	29	49	Popular Headphones

DECODING SHEET FOR COMPARISON TAPE D

Oscar Peterson Trio "The Smudge" -- 7th Chorus

COMPARISON NUMBER	FIDELITY RATING UL GE TL			COMMENTS (OPTIONAL)
100	13	22	15	Pocket Ratio
101	62	71	72	OTE Hearing Aids
102	61	42	53	Popular Headphones
103	38	38	48	Speech Audiometer
104	69	56	87	Monitor Speakers
105	71	69	86	ITE Hearing Aids
106	40	39	55	Discount Stereo

Male Voice "Joe took father's . . ."

110	23	27	18	Pocket Radio
111	81	86	93	Monitor Speakers
112	87	93	95	OTE Hearing Aids
113	87	90	97	ITE Hearing Aids
114	56	41	56	Popular Headphones
115	58	52	69	Discount Stereo
116	53	42	53	Speech Audiometer

N.Y. Philharmonic Beethoven Violin Concerto in D f Passage

120	11	13	11	Pocket Radio
121	67	59	68	Popular Headphones
122	66	51	67	Discount Stereo
123	83	84	88	OTE Hearing Aids
124	77	56	75	Speech Audiometers
125	75	73	80	Monitor Speakers
126	84	83	93	ITE Hearing Aids
(extra) 127	74	77	86	Monitor Speakers vs AR3a's after supertweeter replaced on right AR3a

VITA

The writer recently enjoyed the honor of election to Fellow of
the Audio Engineering Society. Following the adage that "less is
more", the writer's vita is reproduced below in the summarized form
which accompanied that presentation.

> Mead C. Killion was born in Woodstock, Illinois in 1939.
> After receiving his B.A. degree in mathematics from
> Indiana's Wabash College in 1961, Mr. Killion went to work
> for Industrial Research Products, Inc. in Elk Grove Vil-
> lage, Illinois, where he is now senior engineer. He later
> earned his Masters from the Illinois Institute of Tech-
> nology in Chicago (1970). Killion's work at IRPI has
> involved the design of electroacoustical transducers and
> instrumentation, leading to significant contributions to
> the development of, and literature on, miniature microphone
> cartridges for hearing aids. He has presented a number
> of papers before the Audio Engineering Society and the
> Acoustical Society of America, and is coauthor of several
> patents in the miniature transducer field. He is a member
> of the Audio Engineering Society, Acoustical Society of
> America, Institute of Electrical and Electronic Engineers,
> and is active in professional society work notably with
> the Chicago Acoustical and Audio Group, of which he is
> past-president and where he also serves on the executive
> board.

Most of the writer's past research has been in the same general
area as the present dissertation. Thus, most of his published papers
are listed in the References (q.v.), where also co-authored papers
(with Carlson; Halperin, et al.; Knowles; and Studebaker) are to be
found.