

METHODS OF GAUGING THE RADIO AUDIENCE IN THE
LIGHT OF OBSERVED LISTENING BEHAVIOR

By
Martin D. Kaplan

Thesis submitted to the Faculty of the Graduate School
of the University of Maryland in partial
fulfillment of the requirements for the
degree of Doctor of Philosophy

1947

UMI Number: DP70124

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI DP70124

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

ACKNOWLEDGMENTS

The writer is deeply indebted to Dr. John G. Jenkins, the director of this study, for his stimulating advice and assistance, and to Dr. Ray C. Hackman for his many helpful suggestions during all phases of the study. The writer is indebted also to those other members of the staff of the Department of Psychology who were responsible for securing the cooperation of student observers and interviewers. And, of course, the observers and interviewers themselves deserve more than a mere verbal acknowledgment of gratitude.

J. G. Jenkins

TABLE OF CONTENTS

CHAPTER	PAGE
I BACKGROUND	1
Brief History of the Development of Radio	
Audience Research	1
The Mail Questionnaire	4
The Personal Interview	9
The Telephone Day-Part Recall Method	13
Coincidental Telephone Method	18
C.A.S. Coincidental Telephone Method	21
A. C. Neilson Audimeter Technique	26
The Program Analyzer Technique	29
Schwerin - NBC Technique	33
Summary	34
Statement of the Problem	35
II PROCEDURE	37
The Observation Method	37
The Coincidental Telephone Method	41
III RESULTS AND DISCUSSION	45
The Observation Method	45
The Coincidental Telephone Interview Method	68
IV SUMMARY AND CONCLUSIONS	79
SELECTED BIBLIOGRAPHY	85
APPENDIX	87

TABLE OF TABLES

TABLE		PAGE
I	Distribution of Observers and Observees by Programs	38
II	Percentages of Time Spent in the Three Degrees of Listening by Programs	45
III	Percentage of Time Spent in the Three Degrees of Listening by Program Subgroups	52
IV	Coefficients of Determination	56
V-A	Percentage of Observed Minutes of Each Program During Which Each Activity Occurred	61
V-B	Probabilities of the Significance of the Differences in the Percentage of Observed Minutes During Which Each Activity Occurred	62
VI	Percentage of Each Activity Occurring During Attentive Listening, Casual Listening, and Non-Listening	63
VII	Percentage of Male-Observee Minutes and Female-Observee Minutes During Which Each Activity Occurred and the Probability that the Sex Differences are due to Chance	65
VIII	Distributions of the Number of Observees Conversing Varying Numbers of Minutes for the "Commercial-Minutes" Group and the "Non-Commercial-Minutes" Group	67
IX	Outcome of 1020 Coincidental Interviews	69
X	Comparison of National Cooperatings with Cooperatings Obtained in this Study	70
XI	Responses to Question 3 of the Coincidental Interview	71
XII	Answers to Question 3 by Sex of Amos n' Andy Respondents	72
XIII	Frequency of Amos n' Andy Listeners' Responses to Question 3 by Time Periods	73
XIV	Frequency of Responses to Question 3 by Amos n' Andy Subgroups	73
XV	Amos n' Andy Listeners' Responses to Question 4	74
XVI	Amos n' Andy Listeners' Responses to Question 5	75
XVII	Vox Pop Listeners' Responses to Question 4	76
XVIII	Vox Pop Listeners' Responses to Question 5	77

TABLE OF FIGURES

FIGURE		PAGE
1	Profile of a Jack Benny Program	48
2	Profile of an Amos n' Andy Program	49
3	Profiles of Subgroups of an Amos n' Andy Program	53
4	Profiles of Subgroups of a Jack Benny Program	54

CHAPTER I

BACKGROUND

Brief History of the Development of Radio Audience Research

In the fall of 1922, a pioneering advertiser paid Station WJAY in New York the sum of \$100 to sponsor a ten minute radio program. Radio advertising was born. A quarter of a century later advertisers were paying several hundred dollars a minute to reach their buying public. In 1946, they spent about 385 million dollars to talk prospective consumers into buying their products.

Those who spent all this money wanted to know as much as possible about the radio audience. They realized that the value of radio to them was a direct function of the number of potential customers it reached. These advertisers demanded measures of the size and nature of the listening audience-- measures comparable to those developed for the printed page by the Audit Bureau of Circulation. The stage was set for radio audience research. How many people own radio sets? Where do these people live? How many of them listen to a particular program? Does listening to a program affect the buying habits of those who listen? Questions like these challenged the many investigators attracted to the new field of radio audience measurement.

Archibald H. Crossley, a pioneer in radio audience measurement, envisioned three primary objectives for radio audience

research (5), namely, to determine the

1. size of the audience,
2. likes and dislikes of the audience.
3. influence of radio on the buying habits of the audience.

Radio researchers have spent most of their time measuring the size of the audience. They have spent less time gauging the selling power of radio, and even less finding out what people like and dislike about radio programs.

Early investigators confined their efforts largely to the analysis of unsolicited audience mail. It soon became apparent, however, that hero-worshippers, invalids, lonely people, and the very young and the very old contributed a disproportionate share of these letters. Aware of the dangers of predicting from such biased samples, radio research people turned over the fan mail to the sponsors and the talent, and they themselves sought more objective ways of answering the question, "How many people listened to program X last night?"

They turned to the mail questionnaire and the personal interview. During the late twenties, these were the most commonly used techniques for determining the size and composition of the radio audience. These methods too, however, had serious deficiencies.

The first major advance in the measurement of the size of the radio audience came in 1929 when Crossley Inc., in a study for the Eastman Kodak Company, called a random sample of telephone subscribers and asked what programs had been listened to on the previous day. This appears to have been the first detailed study of the audience of a particular program. In

quick succession, Crossley Inc. made similar studies for many more advertisers. All these studies were based on the telephone recall method. Crossley estimated in 1929 that from 80% to 85% of all radio homes could be reached by telephone, and therefore, that the bias resulting from the exclusive use of a telephone sample was not too serious. In contrast, studies in 1939 indicated that only about 45% of the radio homes could be reached by telephone - 55% of the homes with radios did not have telephones. The bias thus introduced was of such magnitude as to require either a large corrective factor in estimating the size of an audience or the use of extreme caution in generalizing beyond the telephone sample.

In 1929, Crossley Inc. proposed an operating plan for a continuing study of the radio audience to be financed by advertisers. The cooperative analysis of Broadcasting was thus set up and field operations were begun in March 1930. Crossley Inc. was retained to carry out the technical research. Four years later C.A.B. became a cooperative, non-profit organization under the direction of a committee composed of three advertising men appointed by the Association of National Advertisers and two agency men appointed by the American Association of Advertising Agencies, a total of five members. Until its dissolution in the summer of 1945, C.A.B. served as perhaps the most important radio research operation.

A second major stride was made in 1930 when Pauline Arnold introduced and promoted the coincidental telephone technique. C. E. Hooper has developed this technique to the point where it is now the most widely accepted method of determining the size

and make-up of the radio audience. In essence, the coincidental method consists in asking the question, "What program are you listening to now?", instead of the G.A.S. recall question, "What program did you listen to _____?"

A third event of considerable consequence in radio audience measurement was Frank Stanton's experimentation in 1933 and later with a device to record mechanically the exact time during which a radio set was turned on. Several other investigators working at about the same time as Stanton, developed similar instruments, the most notable of which was the one constructed by Elder and Woodruff at M.I.T. A. C. Neilson purchased the rights to this device, named it the Audimeter, and spent seven years perfecting it. The Audimeter is now used on a commercial basis to study the habits of the radio audience.

A fourth and recent advance was the development of the Program Analyzer by Lazarsfeld and Stanton in 1937. The Program Analyzer attempts to measure the likes and dislikes of listeners for each part of a radio program rather than the size of the listening audience. It represented the first real attack on the second of Crossley's three objectives of radio research, namely, the determination of the likes and dislikes of the people who listen to radio. At the present time, the Columbia Broadcasting Company, the McCann-Erickson Company, and the Office of Radio Research at Columbia University are engaged in Program Analyzer research.

The Mail Questionnaire

The mail questionnaire was used extensively, particularly

in the early years of radio, to study the size, distribution, composition, and the opinions of the radio audience. It was used also to determine the effectiveness of radio in accomplishing its objectives.

The method has been used to determine what brands of products were used in radio and non-radio homes (3); whether articles were purchased as a result of listening to the radio (23); whether a program was popular enough to sponsor (8). In studies for the Columbia Broadcasting Company between 1929 and 1933, the Price, Waterhouse Co. used the mail ballot to determine station preferences and listening habits (4). In 1933 Kirkpatrick reported the use of the mail questionnaire in a sample of Minneapolis homes to determine the volume and trends in radio listening, program preferences, reactions to advertisements, and the effect of radio on recreation habits (14).

Samuel Barton's Industrial Survey Co. employed the mail questionnaire to measure general listening behavior on a national scale. Until it disbanded in 1946, this company conducted many mail panel studies using a technique they called the Controlled Mail Method, a method first tried on a national sample in the spring of 1940 in a CBS county-by county survey. Two years later, CBS and the Industrial Surveys Co. cooperated in a Controlled Mail Method study in Washington, D.C. In 1943, Industrial Surveys Co. began its continuing study of radio audiences based on this technique and presented the results in monthly reports known as the "Radio Panel." In the following year the "Radio Panel" was put on a weekly basis.

The National Broadcasting Company has also conducted national

surveys using the mail questionnaire. In one such study (10), postcard ballots were sent to about three million people located in all counties in the United States. Each person was asked to name the radio stations he listened to regularly and the one he listened to most. About ten percent of the ballots were returned.

Crossley reported mail studies by Starch and Elder which showed that purchases of certain radio advertised products were more frequent in listening homes than in non-listening homes (6).

A number of investigators have studied technical problems associated with the mail questionnaire. Stanton, for example, reported a survey in which he used the mail questionnaire to find out about the use of radio-classroom equipment. He sent out a follow-up questionnaire to induce cooperation from those who did not answer the original ballot. Stanton found that more people who returned the original ballot owned radio-classroom equipment than did those who returned the second ballot (21).

Suchman and McCandless studied factors influencing the return of mail questionnaires (22). To one group of people they sent an original and a follow-up questionnaire, and then telephoned a random sample of those who did not reply to either questionnaire. To another group, they sent an original questionnaire, a follow-up questionnaire, and an abbreviated second follow-up questionnaire. Results indicated that greater returns were received from people interested in the subject of the questionnaire and from better educated people.

R. H. Ford and J. A. Clausen reported returns as high as 90% when air-mail special delivery stamps were used with the follow-up ballots (9). The same investigators also reported that

it was possible to extrapolate about the residual group if at least three follow-ups were used. Lazarsfeld confirmed this finding on the basis of personal interviews conducted with residual groups¹.

Lazarsfeld (15) compared the mail questionnaire with the personal interview as regards answering the question, "which stations do you and your family listen to regularly?" One member of a family was sent a mail questionnaire and the rest of the family were personally interviewed. He found enough agreement between the two methods to conclude that the bias of the mail questionnaire was small enough so that the method was useful in determining radio network preferences.

Curtis (7) reported a test-retest reliability study of a mail questionnaire designed to determine the degree of listening to 45 specific programs. Respondents were instructed to indicate for each program whether they listened to it regularly, sometimes, or never. The second ballot was obtained on the pretense that a question has been inadvertently omitted from the original questionnaire. Seventy percent of the respondents repeated their original responses and only two percent reversed their original responses.

Barton (1) used the following basic aids to eliminate some of the shortcomings of the mail questionnaire:

¹This finding was reported while commenting about the Ford, R. H. and Clausen, J. A. paper at the meetings of the American Statistical Association in Atlantic City, New Jersey on January 24, 1947.

- *1. Predetermination of the sample based on random or quota selection within the population strata.
2. Permanent identification of the sample selected by means of a control or serial number indicating stratum, block and cell number, and group sequence.
3. Compensation of the respondent, either in advance of work completed or upon return of the desired data depending on the nature of the study.
4. Repetition of contact with variation of incentive or compensation to those members of the sample who had not returned a previous mailing.
5. Double sampling or subsampling of those members of the sample not responding after completion of the series of repetitive mailings. The subsampling of non-respondents being conducted by direct investigator contact".

Applying these principles, Barton obtained the following types of audience data from his mail panel surveys: total audience listening to a specific program; net weekly, monthly, and 13 week audiences; audience turnover; duplication of listening between two or more same-sponsor programs or between two or more competitive programs; audience source and audience dispersion; audience composition by age, sex, geographic location, community-size, economic class, educational level and audience listening-purchasing relationships.

Experience with the mail questionnaire has suggested the following advantages and disadvantages:

Advantages of the Mail Questionnaire.

1. The sample can be controlled insofar as the investigator has complete and adequate data about the population.
2. A wide geographic area can be sampled.
3. Questions can be standardized physically, though this does not insure that they will be standardized psychologically.

4. The questionnaire can be long.
5. The respondent is free to answer the questions at his convenience.
6. Responses by more than one member of a family can be obtained.
7. The bias of an interviewer is absent.

Disadvantages of the Mail Questionnaire.

1. Complete and adequate data about a population may be difficult to obtain.
2. The number of ballots returned may be quite small unless costly follow-up procedures are followed.
3. The absence of an interviewer makes it impossible for respondents to have misunderstood questions clarified.
4. Several weeks or months may elapse before the ballots of all cooperating respondents are returned.
5. The time lapse between the broadcast of a program and the receipt of a mail questionnaire may result in a significant memory loss.
6. There is no control over a respondent as he answers the questions. Some people will copy program names out of the newspaper, etc.
7. Only literate respondents may be sampled.

In summary, it may be said that in spite of certain objections, the mail questionnaire can be used profitably provided that literacy is not a factor, and provided that returns can be increased in number and representativeness by adequate follow-up procedures.

The Personal Interview

The personal interview technique has been used for essentially the same purposes as the mail questionnaire, namely to determine the size, distribution, composition, behavior, and

opinions of the radio audience. Personal interviews have been used on a coincidental basis but for the most part they have been a recall method.

Two types of recall interviews have been employed - the aided recall and the unaided recall. The aided recall has sometimes been called the printed roster interview because the interviewer carries a printed list of programs.

The Pulse Inc. has been one of the principal exponents of the printed roster survey. Headed by Dr. Sidney Roslow, this organization entered the field of radio audience measurement in 1941, and has confined its activities largely to the New York-Philadelphia area. Once a month it publishes the results of a survey of the listening audience of the Metropolitan New York area - the Pulse of New York.

The Pulse Inc. divides each interviewing day into several periods. For each period a roster is prepared which shows the programs broadcast during that period. Interviewing for each period is conducted immediately after the period ends except that interviewing for the night period is carried on the following morning.

The Roslow interviewer first obtains from the respondent a record of coincidental listening. The interviewer asks or observes whether the radio is on and to which program it is tuned. Then, by leading the respondent to reconstruct his activities, the interviewer determines whether the radio had been listened to at any time during the period under study. After determining at what times the radio had been on, the interviewer shows the printed roster and asks which specific programs had

been listened to by the respondent or by any other members of the family.

Using this basic technique, the Pulse Inc. obtains the following kinds of information about the radio audience (19):

1. General listening and specific program listening by age, sex, family status, socio-economic status, educational status, and occupational status.
2. Listening behavior before and after a specific program - adjacent listening behavior.
3. Duplication in listening - the degree to which people listen to programs of competitive sponsors or to more than one program of the same sponsor.
4. Listening - purchasing relationships.
5. Likes and dislikes for particular aspects of programs.

The personal interview has also been used in national radio audience surveys. In November 1945, at the request of the National Association of Broadcasters, the National Opinion Research Center at the University of Denver conducted 3243 personal interviews in an effort to "assess the strengths and weaknesses of the radio industry, to ascertain where radio stands with the public, in order to blueprint a sound plan of action for the future of broadcasting" (16). Although interviewers in this study encountered a considerable amount of specific criticisms of radio, results in general indicated that "the large majority of people in this country are pleased with radio as it is".

The following advantages and disadvantages are common to most personal interview surveys:

Advantages of the Personal Interview.

1. Both telephone and non-telephone subscribers can be reached. Keyrowitz and Fiske reported a study which revealed that families of lower economic levels listened to more small, non-network stations, thus lending further credence to the argument that listening in telephone homes may be quite different from listening in non-telephone homes (17).
2. Better rapport can be established between an interviewer and respondent in a face-to-face situation.
3. A psychological rather than a physical record of listening may be obtained. To illustrate this, Roslow said that "For a period of perhaps five minutes, while the station announcement and terminal commercials are being made, the respondent is psychologically 'deaf', and cannot be considered a true listener until the program proper that he tuned in comes on the air" (19). This statement implied that in a personal interview situation, a respondent could report such moments of psychological 'deafness'. Studies reported by Chappell (2) would suggest that the ability to recall such behavior might be limited and rare.
4. A record of continuous listening over a period of several hours may be obtained. This ensures a balanced sample from quarter hour to quarter hour.
5. Information on more than one program may be obtained from each respondent thus reducing the cost per unit of information. Roslow estimated that 2400 printed roster interviews could obtain as much information as 201,600 coincidental telephone interviews (19).
6. The interview may be long and thorough.
7. The interviewer may make fairly accurate judgments about the composition of the audience (age, sex, economic status, etc.).
8. The respondent is given sufficient time and information to orient himself to the subject.
9. Visual material may be used.

Disadvantages of the Personal Interview.

1. The cost of personal interviewing is usually high.
2. Programs easy to identify may receive bloated ratings.

3. Some respondents report hearing programs not broadcast during the reported hours of set use.
4. It may be difficult to obtain complete roster information about all programs that may be heard in a locality.
5. Surveys are slow because of the time consumed in making calls.
6. The interviewer may bias results unless adequate precautions are taken.
7. The 'not at home' group presents sampling difficulties.

No other method now available provides as complete and detailed information about the listening audience. Because of this unique contribution, the personal interview method will always merit serious consideration. Perhaps its greatest usefulness will be in suggesting theories and leads which may then be checked with other techniques.

The Telephone Day-Part Recall Method

In the field of radio audience measurement the telephone day-part recall method and the Cooperative Analysis of Broadcasting are almost synonymous since the Cooperative Analysis of Broadcasting used this method from the time of its organization in 1930 until 1943. During these years Crossley Inc., an independent research organization, carried out the technical research of C.A.B.

The name telephone day-part recall was derived from the C.A.B. procedure of dividing the broadcasting day into parts and interviewing respondents over the telephone regarding their listening behavior during each of these parts of the day. Originally, C.A.B. divided the day into four-hour periods

beginning at 8:00 A.M. Immediately after a period ended, interviews for that period were begun. Interviewing for the night period was carried out the following morning.

In 1940 C.A.B. reduced the periods to two hours in duration and increased the number of periods to eight. Two years later, C.A.B. resorted to an overlapping method of interviewing. Although they still asked respondents to recall their listening behavior during the previous two hour period, calling was begun a half hour after the period began. The result was that some calls were made immediately after a program ended and others from a half hour to two hours after a broadcast.

Calls were made in thirty three cities from coast-to-coast, in accordance with the distribution of radio sets. To counteract the bias of a random telephone sample, C.A.B. apportioned their calls among four economic groups. Economic level was determined by locating addresses taken from the telephone books on city maps which had been blocked according to these levels.

The C.A.B. interviewing procedure was as follows. First, the interviewer asked if the radio was in use at any time during the two hour period under study. If the answer was "Yes," the interviewer then asked at what times the set had been on and to which programs it had been tuned. If the respondents gave any description of a program which made identification possible, the program was considered as having been heard. Evidence of listening was based on the respondents ability to name either the program, the talent, the sponsor, or the station.

C.A.B. computed a program rating by dividing the number of listeners to a program by the total number of people interviewed.

Surveys were made during alternate weeks and ratings were sent to subscribers twice a month.

The C.A.B. recall ratings and the Hooper coincidental ratings for the same programs did not agree - C.A.B. ratings were consistently higher. This difference and the keen competition between the C.A.B. and the Hooper organizations led to a considerable amount of research designed to explain why the ratings were different.

M. N. Chappell, then psychological consultant to the Hooper organization, reported a series of studies which pointed out some of the shortcomings of the C.A.B. method (2). These studies were based on about six million coincidental calls made during a two year period and the concurrent day-part recall interviews made during the same two year periods. The results are presented below.

Memory Loss in Recall Method.

Since the C.A.B. method was based on the ability of the respondent to remember, its ratings were influenced by factors which affect the memory for a specific program. The following factors were shown to be operative:

a) Influence of Age of the Program.

One hundred and six evening programs were divided into three groups according to the length of time they had been broadcasting. Comparison of coincidental and recall ratings of each group showed that programs over two years in age were remembered about 18% better than programs less than one year in age.

b) Influence of the Length of the Program.

Fifty four evening programs (over two years in age) with lengths from 15 minutes to an hour were studied and it was found that one hour programs were remembered about 52% better than programs of the same

age but only 15 minutes long.

c) Influence of Program Popularity.

Forty two evening programs over two years in age and all a half hour in length were compared. Programs receiving coincidental ratings of 15.0 or more were remembered about 11% better than those receiving coincidental ratings of less than 11.0.

d) Influence of Type of Program.

A study of eighty two evening programs without regard to age, length, or popularity (and thus, some caution must be exercised in interpreting these conclusions) revealed that the memory for variety programs is greatest. Next in order are concert music, plays, continuity drama, quiz, popular music, miscellaneous, and news programs.

e) Influence of Network.

A study of eighty two evening programs showed that memory for NBC Red (now NBC) and CBS programs was greater than memory for NBC Blue (now ABC) and MBS programs. Part of these differences is attributable to the fact that NBC Red and CBS carried older, more popular types of programs than did NBC Blue and MBS.

Chappell and Hooper also pointed out the shifting base of the recall method (2). The base of the recall rating was the number of people interviewed - the number who answered the telephone. The 'not at home' group were omitted from the sample. It may be reasonable to assume that the 'not at home' individual was not listening to the radio at the time of the call, although he might have been listening to an automobile radio or to a radio in another location. It is certainly not reasonable, however, to assume that the individual did not listen to the radio at some time during the two hour period covered by a recall interview.

The C.A.S. procedure of omitting the 'not at home' group

would have been less objectionable if it had not been shown that this group varied in size from time to time and from place to place. For example, under normal conditions there is no one at home in about 50% more homes in July and August than in January and February. This is further confirmed by a comparison of coincidental and recall ratings during the winter and summer - recall ratings were about 9% higher than coincidental ratings in the winter and about 31% higher in the summer. The differences in 'not at home' groups in different parts of the country also results in discrepancies between the two rating systems. In general, as the 'not at home' group increased in size, the difference between the ratings increased.

The time elapsing between a broadcast and an interview has also been shown to affect ratings. In an interview covering a two hour period, respondents may report that they had listened to programs which were broadcast from a half hour to two hours before the interview. Chappell reported that the size of a recall rating was inversely related to the amount of time elapsing between the broadcast and the call. This difficulty was overcome to some extent when C.A.B. began its overlapping system of interviewing.

In spite of all these objections to the C.A.B. recall method, it was the most widely used method of audience rating for many years. It was not until 1942 that C.A.B. was ready to accept the cold facts. In that year, the recall method was abandoned, and in its place C.A.B. adopted the coincidental method which Hooper had used so successfully since 1934.

The C.A.B. coincidental method differed from the Hooper

method in what seemed at first to be minor technical points. In time, however, these points proved to be major ones and C.A.B. and Hooper were still unable to agree on ratings for the same program. In the summer of 1946, C.A.B. ceased operations in the field of radio audience measurement.

Coincidental Telephone Method

The coincidental telephone method is the most widely accepted technique for determining how many people listen to a radio program. The method has long been associated with the name G. E. Hooper. Although Pauline Arnold was the first to promote the method early in 1930, it was Hooper - first as a member of the Hooper-Holmes Bureau and later under the name of G. E. Hooper Inc. - who developed it to its present status.

The Hooper interviewer calls people on the telephone while a program is on the air and asks the following questions: (2)
(12).

1. Were you listening to your radio just now?
2. To what program were you listening, please?
3. Over what station is that program coming?
4. What advertiser puts on that program?
5. Please tell me how many men, women, and children including yourself were listening to the radio when the telephone rang?
6. What is the occupation of the head of your household?

Any one of the following situations may be encountered when the telephone rings.

1. The operator may report that the telephone has been disconnected in which case the call is dropped from the sample.
2. There may be no answer. If the telephone rings six

times without an answer, it is assumed that no one is 'at home and awake', and the home is classified as a non-listening home.

3. A busy signal may be obtained, in which case it is assumed that someone is 'at home and awake'. It is also assumed that the listening behavior in such homes is the same as the listening behavior in the 'at home and awake' homes in which someone answers the telephone. Therefore, they are retained in the sample and prorated according to what is found in the other 'at home and awake' homes.
4. The telephone may be answered, in which case question 1 above is asked. If the respondent says that he was not listening to the radio when the phone rang, the call is terminated. If he was listening, the remaining questions are asked. From the answers to questions 2, 3, and 4 above, it is determined whether the respondent was listening. Any identification of the program, sponsor, talent, or station is considered sufficient evidence to classify the respondent as a listener to a particular program.

The Hooperating for a program is obtained by dividing the number of listening homes by the total number of homes called. It is the percentage of homes called in which the respondents were listening to the program.

The regular network Hooperatings are based on calls made in thirty three cities in the United States. These cities were selected because each has local outlets for all four major networks and thus, presents relatively uniform conditions of network competition.

The following size samples are obtained for the regular network ratings:

15 minute program broadcast once a week
630 homes called

30 minute program broadcast once a week
1350 homes called

60 minute program broadcast once a week
2700 homes called

15 minute program broadcast five times a week
4050 homes called

In addition to the regular network ratings, C. E. Hooper Inc. computes an available audience rating. The available audience rating is the ratio of the number of homes listening to a program to the number of 'at home and awake' homes. It is intended to reflect the degree to which a program can pull an audience from the total number of homes available to listen - it is a measure of the attraction value of a program. Two programs broadcast at different times may have the same regular network rating but because of a difference in the size of the available audience, one may have a higher attraction value. Here is an illustration:

	Total Homes Called	At Home and Awake	Not Home	Listening to Program	Network Rating	Available Audience Rating
Program A Sunday	1000	700	300	150	15.0	20.1
Program B Monday	1000	800	200	150	15.0	18.8

Other types of information may also be obtained from a coincidental telephone survey. Audience composition may be broken down by age, sex, and occupation of the head of the household. The ability of listeners to identify sponsors may be determined. By calculating ratings at regular intervals during a program, it is possible to ascertain whether a program is an audience builder or an audience dissipater. All these analyses

are made available to subscribers to Hooper ratings.

In spite of its widespread acceptance, the coincidental telephone rating has some serious limitations. The most obvious is that it is limited to telephone homes. The rating cannot be projected to the nation-wide listening audience. It serves only as a measure of comparative network program popularity among telephone subscribers in the thirty three cities included in the Hooper sample. Because of its very nature, the method can never yield anything but a measure of radio listening in telephone homes. In contrast, the A. C. Nielsen audimeter technique - described in a later section - may be made to represent nation-wide listening if subscribers are willing to spend enough money to make such a venture profitable and if it can be established that radio set tuning and radio audience listening are essentially the same. No expenditure of money, however, can ever make the coincidental telephone rating representative of non-telephone subscribers. The Hooper organization realizes the limitations of its method and is planning a national audience rating but it will be based on data from family diaries rather than from coincidental telephone calls.

O.A.B. Coincidental Telephone Method

In 1942, O.A.B. adopted the basic Hooper method for the determination of its program ratings. The O.A.B. coincidental technique, however, differed from the Hooper technique in several respects. Before 1942, while O.A.B. was using the day-part recall method, differences between O.A.B. ratings and Hooper ratings were commonly said to be due to the widely different

techniques involved. After 1942, however, differences between these two ratings could not be dismissed so easily. To the dismay of subscribers, striking differences still occurred. Which rating should they accept?

Zeisel found that the two ratings correlated 0.97 but that the C.A.B. ratings were on the average 17% lower than Hooper ratings for the same programs (24). This consistent variation may be explained in terms of some of the procedural differences between the two techniques.

Procedural Differences between the C.A.B. and Hooper Coincidental Techniques as of 1946.

1. Number of cities used in the samples.

Hooper ratings were based on interviews in 33 cities, each of which could clearly receive all major network programs.

C.A.B. ratings were based on interviews in 81 cities. Actually, however, a full network hookup rarely reached more than 65 of these cities so that C.A.B., in practice, used only about twice as many cities in its sample as did Hooper.

2. Number of calls.

For a half hour, evening, full network program, Hooper sampled 1350 homes; C.A.B., 1780.

3. Interviewing dates.

Hooper interviewed during the seven day periods beginning with the first and fifteenth of each month (for evening programs).

C.A.B. interviewed during the seven day periods beginning with the first and third Saturdays of each month (for evening programs).

4. Interviewing minutes.

Hooper interviews were conducted from the third through the fifteenth minute of each quarter hour broadcast.

C.A.B. interviews were conducted from the second through the fourteenth minute of each quarter hour broadcast.

5. Rating of programs broadcast after 10:30 P.M. in the Eastern time zone.

During one interviewing week of the month, Hooper interviewers made calls after 10:30 P.M. in the eastern time zone in spite of the lateness of the hour. The ratio of listening in the east to listening in the rest of country was determined and it was assumed that this ratio remained the same during the second interviewing week of the month when the 10:30 P.M. deadline was not broken. The validity of this assumption can easily be tested though there is no report of such a test in the literature.

C.A.B. conducted recall interviews in the eastern and central time zones the following morning, in addition to the routine coincidental interviews in the central zone. The coincidental rating for the eastern time zone was computed from the following formula:

$$\frac{CR \text{ (east)}}{CR \text{ (central)}} = \frac{RR \text{ (east)}}{RR \text{ (central)}}$$

or

$$CR \text{ (east)} = \frac{RR \text{ (east)} \times CR \text{ (central)}}{RR \text{ (central)}}$$

CR - coincidental rating
RR - recall rating

This formula assumed that the ratio of the coincidental and recall ratings for a program in the central zone was equal to the ratio of the coincidental and recall ratings for the same program in the eastern zone. In the absence of experimental evidence, the validity of this assumption is questionable. Indeed, the bulk of psychological research on forgetting would seem to invalidate such an assumption. A program heard at 10:30 P.M. in New York would have been heard at 8:30 P.M. in Chicago. If forgetting "is not so much a matter of the decay of old impressions and associations as it is a matter of the interference, inhibition, or obliteration of the old by the new" (13), then it is reasonable to believe that a Chicagoan would be

more likely to forget his listening behavior of 8:30 the previous evening than would a New Yorker his listening behavior of 10:30 the previous evening. If so, then the recall-coincidental ratios would not be equal from one time zone to another.

6. Wording of the opening question of the interview.

Hooper began his interviews with the question, "Were you listening to the radio just now?"

C.A.B. began with the query, "Will you please tell me what program you were listening to on the radio when the telephone rang?"

It is true that the C.A.B. question is a leading one but this would tend to bloat CAB ratings and therefore this difference in procedure cannot explain even in part why C.A.B. ratings were found to be 17% lower than Hooperatings.

7. Treatment of Busy lines.

Hooper retained these in the sample and prorated their program listening according to the rest of the 'at home' group.

C.A.B. discarded these cases from the sample on the theory that there were too many party lines to assure that a "Busy signal meant that the person called was at home." Some of the 17% variation between the two ratings can be accounted for by this technical difference in procedure.

8. Treatment of listeners who do not identify a program.

A certain number of people reported that they were listening to the radio but either could not or would not identify the program to which their radio was tuned. Hooper retained these cases as sets-in-use and redistributed them among the programs in the same proportions as the rest of the sample.

C.A.B. retained these cases as sets-in-use but did not assign them to any particular program. This was obviously underestimating the size of program audiences and this procedural difference too can account for a certain amount of the variation between the two ratings.

9. Determination of 'party not at home.'

Hooper interviewers allowed six rings before assuming that the party was not at home.

C.A.B. allowed only four rings before making the assumption that no one was at home.

In a large scale study by Zeisel (24), it was found that "of all parties who answer before the end of the tenth ring, only 96.0% will have answered at the end of the sixth ring (Hooper), and only 86.4% will have answered at the end of the fourth ring (C.A.B.).

In summary, though C.A.B. ratings correlated 0.97 with Hooper ratings, the former were on the average 17% lower than the latter. A major part of this difference can probably be accounted for by the fact that Hooper allowed six rings and C.A.B. allowed only four rings before deciding that a party was not at home. Additional variation between the ratings can be explained by procedural differences in the handling of non-identifiers and busy lines.

Zeisel reported a flaw common to both C.A.B. and Hooper which resulted in a serious underestimate of the audience size by both ratings (24). When a respondent answered "No" to the question "Were you listening to the radio just now?" (or to the opening C.A.B. question), Zeisel asked if anyone else in the house was listening. He found that 9 out of 100 houses were actually listening, though both C.A.B. and Hooper would have classified them as not listening.

As mentioned earlier, C.A.B. ceased operations in the field of radio audience measurement in the middle of 1946. This decision to discontinue was probably due in a large part to the studies of Zeisel which pointed to the conclusion that C.A.B. was underestimating the size of the radio audience by at least 20%. Faced with these facts, advertisers could hardly be

expected to subscribe to C.A.S. reports while Hooper reports were available. The Zeisel experiments will probably also lead to an improvement in the Hooper technique.

A. C. Neilson Audimeter Technique

In 1942, after seven years of experimentation with the Audimeter, the A. C. Neilson Co. made its debut into the highly competitive field of radio audience measurement. Because this organization is the only one now using a mechanical recorder, the A. C. Neilson Audimeter has come to be known as the mechanical recorder in radio audience research.

When installed in the receiver of a radio set, the Audimeter, operating silently and unseen, makes a graphic record of all of the times at which the set is turned on or off, and the stations to which it is tuned at every instant of the day or night for a period of a month. The monthly record is made on a tape about 100 feet long and 3 inches wide and at the end of the month the tape must be removed and replaced.

At the present time, Audimeters are installed in about 1140 homes in the United States (18). Representing 63% of the area of the country, the Neilson sample has been selected with regard to radio set distribution, city size, family size, type of dwelling, number of rooms, education, and income. The representativeness of the sample is under continuous surveillance and when indicated, the necessary changes are made. The annual turnover in the sample approximates twenty percent.

Once a month, Neilson agents visit each home in the sample and conduct pantry checks. At the same time, the used tapes are

recovered and replaced with new ones. For one reason or another, a certain number of the tapes are not recovered and about 10% of those recovered are defective, thus reducing the usable sample to about 1000.

The Neilson Radio Index - NRI - is a monthly publication based on the analysis of about 1000 tapes. The report contains the following kinds of information for each sponsored network program (16):

1. Size of Audience
2. Nature of Listening
3. Quality of Audience
4. Cost of Audience
5. Product Usage by Audience
6. Sales Effectiveness

Advantages of the A. G. Neilson Audimeter Technique.

1. The sample is fixed and therefore its characteristics may be determined quite readily. Fairly accurate observations of these characteristics may be made when the device is installed and checked each time an agent calls to recover a tape.
2. With adequate funds, the size of the sample could be extended to give nationwide coverage. This is not true of a telephone method which, by its very nature, can never be representative of anything but telephone subscribers regardless of how large a sample is used.
3. Memory is not a factor.
4. Audience turnover can be measured effectively.
5. Partial and complete listening can be identified and treated separately.
6. Long term trends can be studied.
7. Minute-by-minute listening activity can be measured.
8. Product usage and sales effectiveness can be determined from pantry studies.

Disadvantages of the A. G. Neilson Audiometer Technique.

1. The device records set usage rather than audience listening. The seriousness of this criticism is yet to be determined by careful experimentation. Although, on the basis of cursory evidence, Neilson reported that there is no "substantial difference between listening and tuning," the question is still open to investigation (18).
2. The presence of the device in the home may in itself affect listening activity.
3. Families permitting installation of the device may not be representative from a listening point of view even though they have been carefully selected according to other factors.
4. The size of the overall sample is so small - about 1000 - that analysis of subgroups may produce unreliable results. Only about 5% of the available sets are usually tuned to the average daytime program. This means that in about 50 homes in the Neilson sample, the radio is tuned to a daytime program. Any attempt to subdivide this group of 50 according to city size, economic status, or type of dwelling would result in a ridiculously small sample.
5. The necessity of recovering tapes from widely dispersed homes results in an extended time lag between the broadcasting of a program and the preparation and distribution of reports to clients.

In fairness to the method, it should be pointed out that the last two criticisms are not directed at the method per se but rather at the present usage of the method. If sufficient funds were available, the size of the overall sample and the frequency with which tapes are recovered could be increased. As a matter of fact, at the present time the A. G. Neilson Co. is experimenting with a new device known as the Instantaneous Audiometer. Located in a home anywhere in the country, the Instantaneous Audiometer will transmit a continuous stream of electrical impulses to the Neilson Building in Chicago showing every movement of the dial made in the home. Fifty such instruments have already

been built. When this device becomes available for wide-scale use, and when the size of the overall sample has been adequately expanded, the Neilson technique will stand or fall on the answer to the question of whether it measures listening or tuning activity. It can be said, at this point, that data to be presented in a later section of this report suggests that listening and tuning activity may be quite different.

The Program Analyzer Technique

The development of the Program Analyzer by Lazarsfeld and Stanton in 1937 was a landmark in radio audience research. It represented the first successful attempt to record and study audience reaction to programs and program components. In their typical 'why' approach, Lazarsfeld and Stanton sought not only a detailed description of audience reactions but also the reasons for the reactions.

The Program Analyzer is essentially an adaptation of the laboratory polygraph. It permanently records, in terms of like, dislike, and indifference, listeners' reactions to every part of a radio program. To indicate approval, the listener presses a green button and keeps it depressed as long as approval continues. To indicate disapproval, he presses a red button and keeps it depressed as long as disapproval continues. To indicate indifference, he refrains from pressing either button. The buttons are connected electrically to a series of pens which rest on a continuously moving tape. When a button is depressed, the appropriate pen makes contact with the tape and makes an imprint as long as the contact is maintained. Separate records are made

for each individual tested.

A typical Program Analyzer session might run this way. Cooperating subjects are gathered together in a studio and instructed in the use and operation of the Program Analyzer. Each individual is provided with a green and a red button. When the group has been put at ease, the test recording is begun. Throughout the program, the listeners indicate their second-by-second reactions to what they hear. A trained interviewer, in an adjoining room with the polygraph, studies the records as they are made and familiarizes himself with the reaction patterns of each listener. After the program, the interviewer supervises the filling out of a questionnaire designed to get at general reactions to the program as a whole. Then the Program Analyzer records are brought in and placed on the wall so that all the listeners can see them. Group interviewing is begun. Questions are directed to individuals as well as to the group. The tape is constantly referred to in an effort to uncover reasons for every depression of a button. Frequently, a portion of the recording is replayed to help the participants recall the reasons for their reactions. Oftentimes, the interviewer poses an individual's reaction to the whole group for a show of hands. All questions and answers are recorded by a stenographer.

A Program Analyzer session produces three sets of data - Program Analyzer records, questionnaires, and group interview transcriptions. From the Program Analyzer records, a group profile is drawn which shows the group's reaction to each phase of the broadcast. Interview material and questionnaire results are related to the profile in an attempt to uncover the reasons for

peaks of approval and disapproval.

In most Program Analyzer studies, only ten to fifteen participants have been used in any one session. McCann-Erickson Inc., by contrast, has a Program Analyzer which can simultaneously record separate records for fifty listeners. The Columbia Broadcasting System's "Big Annie" can handle one hundred individuals at a time but can produce only a group record.

The reliability of the Program Analyzer has been tested in several ways. Schwerin reported a reliability of .89 based on two groups of 19 individuals who listened and indicated their reactions to a 15 minute news program (20). The correlations between the scores for the even and odd portions of three different programs were .99, .97, and .95 based on 79, 39, and 33 listeners respectively (11). In another study, four matched groups of 45 listeners expressed their reactions to the same program. Correlations of the rank order of program parts for each group with all other groups ranged from .60 to .80 (11).

In the absence of any outside criteria, the validity of the Program Analyzer was checked by having a group of listeners 'Program Analyze' a program which had been modified on the basis of findings in an earlier 'Program Analysis'. In one such check, it was found that an opening portion of a program scored - .40 in the original form and +.57 after modification, a negative score indicating group disapproval and a positive score indicating group approval (11).

In another type of validity check, Program Analyzer scores were compared with expressions of interest in the program as a whole as indicated in a written questionnaire. Results showed

that people very interested in a program gave it higher scores on the Program Analyzer than did people who were interested, indifferent, or bored (11). The regular listeners to a news program averaged + 1.00 on the "Program Analyzer" of that program as compared with +.34 for non-listeners (11).

Advantages of the Program Analyzer Technique.

1. It permits people to record their reactions to a program while they listen to it.
2. It makes possible a detailed study of specific aspects of a program such as the commercial, the announcer, the master of ceremonies, the sound effects, the production technique.
3. It enables audience opinion as well as expert opinion to guide those whose responsibility it is to evaluate the merit of a program.

Disadvantages of the Program Analyzer Technique.

1. The experimental listening situation is by necessity artificial. It is quite unnatural to listen to a program in a studio by invitation and to express, second after second, judgments of like, dislike, and indifference. It cannot be assumed that the same expressions would occur under normal, less critical, home-listening conditions.
2. The terms like, dislike, and indifference may be interpreted differently by different people. Villainous characters in dramas may be disliked because they have been successful in carrying out the impressions which they were designed to create. The group interview has revealed, also, that people with an overall favorable impression and people with an overall unfavorable impression of a program, as determined from a written questionnaire, used the indifference reaction to mean different things. The people who were favorable to a program tended to react to parts they didn't like by failing to press either the green or red button, i.e., with an indifferent reaction. On the other hand, people with an overall unfavorable impression tended to react to parts they liked with an indifferent reaction. The group interview, of course, has the effect of revealing and therefore counteracting some of these discrepancies.

3. In the group interview, some respondents may not be as willing or as able as others to express reasons for likes and dislikes, particularly in the face of reasons already expressed by others who disagree with them.
4. The nature of the equipment and the experimental design make large scale sampling impracticable.

In spite of certain shortcomings, the technique has unquestionable value. Without some satisfactory measure of audience reaction, such as is provided by the Program Analyzer, radio talent and radio producers must grope in the dark in search of ways to bolster sagging cooperatings. Though still an infant, the Program Analyzer can provide some help in discovering, diagnosing, and doctoring the sore thumbs of radio.

Schwerin - NBC Technique

Horace Schwerin, a consultant to the National Broadcasting Company, has developed another method for the measurement of audience likes and dislikes. The method bears a remarkably close resemblance to the Program Analyzer technique.

Cooperating listeners gather in a studio and listen to a transcription of the program being studied. Each participant is given a sheet of paper on which he records expressions of like, indifference, and dislike at the end of each predetermined unit of the program. By appropriate signals, the master of ceremonies indicates that a response unit has ended and that an entry is to be made. After the program has been played and the record sheets have been collected, a group interview is conducted by the master of ceremonies. The audience is prompted to discuss the reasons for their previous expressions and after

adequate debate, they vote approval or disapproval of each unit by a show of hands.

It is anticipated that a mechanical device will be substituted for the written record sheet. When this is done, the method will even more closely resemble the Program Analyzer technique.

Any criticisms which may be leveled against the Program Analyzer technique may also be directed against the Schwerin approach. In addition, the Schwerin method probably interferes more with normal listening than does the Program Analyzer method. Pressing a button when one feels moved to do so would seem to be less disturbing than making a written entry when a master of ceremonies gives the cue.

Judgment as to any unique contributions of the Schwerin technique must be held in abeyance until more results of the use of the method are made available.

SUMMARY

We have looked in some detail at many of the methods currently available for studying the radio audience. We have discussed the mail questionnaire, the personal interview, the day-part recall telephone interview, the coincidental telephone interview, the mechanical recorder, the program analyzer, and the group interview.

We have talked about the good and bad in each of these methods. From what has been said, it should be clear that no one method is perfect - no one method will provide all the answers. The method which one would use in any particular investigation

will depend upon the kind of information desired, the time and money which can be expended, and finally, the use to which the obtained information will be put. A common and highly recommended practice is to use two or more methods in combination. The methods selected for such combinations should compensate for each others limitations insofar as possible.

The measurement of sheer size of the radio audience can be accomplished reasonably accurately by using combinations of existing methods. As these methods are refined and extended, audience size will be measured with even greater assurance of accuracy. The road ahead for research is straight and clear.

On the other hand, the measurement of what people like and dislike about radio programs presents a more serious challenge. The road here is winding and cluttered with obstacles. Not until the road is cleared will a practical and adequate methodology emerge. The Lazarsfeld-Stanton Program Analyser is a first step. In the pages which follow, the writer will describe a few more steps in the determination of what people think about radio programs.

Statement of the Problem

This research accepts as its primary objective the adaptation of two methods - the observation method and the coincidental telephone interview method - to the measurement of audience reaction to radio. With the observation method, we will observe what people actually do while they listen to the radio; with the coincidental telephone interview method, we will ask people directly what they think about a program while they are listening

to it.

A basic assumption underlying this investigation is that the likes and dislikes of the people who listen to radio can best be determined from a study of their normal, home-listening behavior, rather than from a study of what they do and say while listening to a program in the artificial surroundings of a testing studio.

A secondary objective, but one with broad implications, is the establishment through the observation method of a criterion of audience behavior. Both Hooper and Neilsen have implied that their methods measure audience behavior. From Chappell and Hooper, we have the statement that, "The first requirement of any method for measuring radio audience size in any selected population is that it will reflect accurately what people do" (2). Neilsen, in a discussion of what he termed the overemphasis of the difference between listening and tuning, presented cursory evidence to support the conclusion that the difference between the two is "not substantial" (18), and that therefore, his method does measure listening behavior.

If the observation method can furnish a record of what people actually do while they supposedly listen to the radio, then it may serve to provide a criterion against which to validate the Hooper and Neilsen methods, or indeed, any of the other currently available methods which purport to measure some aspect of radio listening behavior.

CHAPTER II

PROCEDURE

The Observation Method

One hundred and ninety eight students in psychology classes at the University of Maryland acted as observers in this study. Those who lived at home observed their families and friends, and those who lived on the campus observed their roommates, sorority sisters, or fraternity brothers. Unlike most observation studies in which the observer is hidden from the view of the observees, the observers in this investigation carried on their work in the same room with the people being observed. This was necessary if we were to be able to observe normal listening behavior in the home situation. The observees did not know, however, that they were being observed. The suspicions of inquisitive observees were easily dispelled by the observers with the statement that they were doing homework, or writing a report, or analysing a radio program for a psychology class. That the observers effectively camouflaged their work is borne out by the fact that only 4 of the 198 - less than 2% - were discovered by the observees. The data of these four observers were discarded along with those of 12 others who obviously failed to follow the directions and 37 others who began observing after the program had been under way.

There remained for analysis, then, the data of 145 observers who observed 254 people for a total of 7620 minutes of radio

'listening'. Table I shows the distribution of these observers and observees by programs.

TABLE I
Distribution of Observers and
Observees by Programs

Program	Date	Observers	Observees		Total
			Male	Female	
Jack Benny	3-9-47	28	21	33	54
Amos n' Andy	3-18-47	34	28	30	58
Duffy's Tavern	3-26-47	17	12	15	27
Frank Sinatra	3-26-47	8	13	2	15
Miscellaneous	3-9-47 to 3-28-47	<u>58</u>	<u>35</u>	<u>64</u>	<u>109</u>
		145	110	144	254

Members of the staff of the Department of Psychology cooperated in trying to motivate the observers by making the task a regular class assignment. All the observers in this study were inexperienced in such work and they depended heavily, therefore, upon the instructions given to them. Here are the printed instructions which each observer received.

INSTRUCTIONS

1. Record all activities of the individuals being observed in the space to the right of the time column. These entries are to be made directly to the right of the appropriate time and at the time the activity is going on.
2. Activities may take the form of reading, writing, smiling, laughing, talking, manipulating the dial, entering or leaving the room, commenting about the program, etc. Record these activities as they occur and in each case indicate which member of the family is involved.
3. Record verbatim any conversation related to the radio. Record other conversation as general conversation.
4. Before you begin to observe, be sure that your watch is synchronized with radio time.

5. Here is an example of what a record of observations might look like. This was part of an actual case.

List Att.	List Cas.	Not List	Time	Observations
	D, M		8:20	Dad picks up magazine and begins to read. Mother sews.
X	O		:21	Mother grins at joke. Dad still reading.
D, M			:22	Dad puts down magazine, looks at radio, and laughs.
		D, M	:23	Mother asks Dad a question about John L. Lewis
		D, M	:24	Conversation about Lewis continues.
X	O		:25	Mother turns dial during commercial but turns it back to Jack Carson after a few seconds.

6. Notice that there are three columns to the left of the time column (see example above). The first is labeled List Att. (Listening Attentively); the next is List Cas. (Listening Casually); the third is Not List (Not Listening).

For each minute of the program you are to indicate your personal estimate of the degree to which each person being observed is listening to the program. In the example above, the observer indicated that both Dad and Mother were listening casually at 8:20. At 8:21 Mother was listening attentively and Dad was listening casually, and so on.

The following criteria will help you decide which categories to use:

List Attentively - an entry here means that you are sure that the person was listening. If the person smiles at a joke or comments about the program or does something else which you can readily observe, you can indicate that he is listening attentively. You may be able to tell that the person is listening from the way he is sitting or from the direction of his glances.

List Casually - an entry in this column means that the person may be listening but you are not sure. If you are in doubt use this column.

Not Listening - an entry here means that you are sure that the person is not listening. If you see that the person is not attending to the radio, use this column. For example the person may leave the room or he may be so deeply engrossed in conversation or other activity.

7. Be sure that the program being listened to is clearly indicated on the record sheet. Whenever the program is changed for any reason, indicate who made the

change and indicate the new program.

5. After your observations are completed, you may wish to make some comments. Use the back of the record sheet for that purpose.

These instructions were read to them in their psychology classes and specific points were clarified and amplified in answer to questions regarding various phases of the task.

In short, during each minute of the program, the observer's task consisted of (1) recording what he saw and heard his observees do and say, and (2) estimating whether his observees were listening attentively, listening casually, or not listening.

Double Observer Reliability.

Six pairs of siblings - all students at the University of Maryland - observed members of their families for from 30 to 60 minutes of radio listening, a total of 735 minutes. Both members of each pair observed the same people at the same time. The writer carefully explained to these observers the necessity of making independent judgments and entries. At least one member of each pair was enrolled in a psychology class at the time and was not, therefore, completely naive about the need for independent observing. There is no reason to believe that any collusion occurred. The observations provided in this double observer study served as the basis for one of the several estimates of reliability reported in the chapter on results.

Observation Method - Hooper Telephone Interview Cross Check.

The reader will recall that the regular network Hooperatings are determined by dividing the number of respondents who said they were listening to a program by the total number of homes called. In addition, gross audience size is estimated from answers to the question, "Please tell me how many men, women, and children, including yourself, were listening to the radio when the telephone rang?"

How well do telephone responses agree with observed listening behavior? To check this, we called the homes of twenty families who were being observed at the time by our student observers and asked respondents the same

questions which a Hooper interviewer would have asked. The observers were unaware of the fact that these calls were in any way connected with the observation project.

The Coincidental Telephone Method

The second major phase of this research was concerned with the adaptation of the coincidental telephone interview to the measurement of audience reaction. Over 150 pretest calls were made in an effort to develop such an interview.

Many different questions were asked in these pretest interviews in search of ones which would induce respondents to verbalize their opinions about a radio program. Here are some which were tried but discarded:

- What do you think about tonight's program?
- What do you like about tonight's program?
- What do you think of the program so far tonight?
- What do you dislike about tonight's program?
- What was the poorest part of tonight's program?
- Is there any way you think tonight's program could have been improved?
- If you had been the producer of tonight's program, which parts would you have left out?
- What part of tonight's program do you particularly like?
- What part of tonight's program do you particularly dislike?

From this pretesting, there evolved finally a five question, minute-and-a-half interview. The first two questions are the standard Hooper questions designed to find out if the respondent was listening to the radio and if so, to which program. The last three questions aim to tap listener reaction. Here are the five questions:

1. Were you listening to the radio just now?
2. To what program were you listening, please?

3. Compared with other _____
 (insert name of program being
 programs, do you think tonight's pro-

 listened to) _____
 gram so far is excellent _____; good _____; or just
 fair _____?
4. What was the worst part of tonight's program?
5. What was the best part of tonight's program?

The complete interview form appears in the Appendix.

Sixty eight students in psychology classes at the University of Maryland served as the telephone interviewers in this study and were each paid one dollar for their services. All the interviewers were inexperienced in such work. Each one made 15 interviews, a total of 1020, between 9:07 and 9:30 P.M. on March 25, 1947. By starting the first interview at 9:07, we allowed the respondents enough time to form some opinion of the program and also allowed the interviewers enough time to complete their interviews by the end of the program.

Below are the detailed, printed instructions which were given to each interviewer:

READ THESE INSTRUCTIONS CAREFULLY

In addition to this instruction sheet, you will find in your envelope 15 interview sheets and 15 index cards. In the upper right hand corner of each index card is a telephone number which you are to call. In all, you will call 15 numbers. There is an interview sheet for each call and one extra for practice.

Before you begin any calling, fill in a telephone number on each interview sheet. Do this in the space labeled 'telephone number called'.

Dial your first number at 9:07 and then make the rest of your calls with as little time interval between calls as possible. All your calls should be completed by 9:30. If you have some left over at 9:30, do not make them.

Indicate on the interview sheets the time each call was begun. Be sure your watch is synchronized with radio time.

If the phone does not answer after it has rung 6 times, hang up and indicate that there was no answer.

If you get a busy signal, hang up and indicate it on the sheet. Put that sheet at the bottom of your pile and call it again after you have called the other numbers. When you do call it again indicate the time and the result. If you get an answer go through with the interview.

Begin each interview with the statement, "This is the Radio Research Bureau calling." Then ask question 1. If the respondent answers 'no' to question 1, say "Thank you" and hang up. Check 'no' on the interview sheet and go on to the next telephone call.

If the respondent answers 'yes' to question 1, ask question 2. He may tell you the station instead of the program. If he does, ask "Do you know what program is on that station?" If he cannot identify either the program or the sponsor, say "Thank you" and hang up. Indicate on the sheet that he did not know the program. Its alright if the respondent asks someone else in his family for help in identifying the program but indicate on the sheet that he got help and go through with the rest of the interview.

If the respondent does tell you the program he is listening to, say "We'd like to ask you a few questions about that program, if you don't mind." Then ask question 3. In asking this question fill in the name of the program he was listening to. If he was listening to Amos and Andy say, "Compared with other Amos and Andy programs, do you think"? The respondent may say that this was the first time he ever heard that program. If he does, repeat the question in these words, "Compared with other radio programs, do you think this program so far is"?

Check whichever answer is given to question 3. Also record verbatim anything else the respondent says.

Then ask question 4. You may have to probe some to get the respondent to answer this question and the next one. If the respondent seems hesitant about answering, ask the question again in slightly different words. Say, "We'd like to know what you think was the worst part of tonight's program." Allow the respondent a few seconds to gather his thoughts. Record verbatim anything he says. These same instructions apply to question 5.

After question 5 has been answered, conclude the interview by thanking the respondent.

Return all interview sheets and index cards. Also return this instruction sheet. If you have made any toll calls, indicate on the back of this sheet the number and cost so that we may reimburse you.

Be courteous to respondents at all times.

The writer discussed these instructions with the interviewers in a group session in an effort to clear up any misunderstood parts. During this discussion, particular emphasis was placed upon the necessity of probing for answers to questions 4 and 5.

In order to maintain comparability with the samples used in the observation studies, the telephone sample used in this phase of the research was selected at random from the University directory - all the numbers called were of families of students. Although it wasn't particularly important what day and hour was selected, the decision to use March 25, 1947 from 9:00 to 9:30 P.M. was influenced by the fact that just one week earlier, during the same half-hour period, 34 students had observed the listening behavior of 58 people during the Asca n' Andy program.

CHAPTER III

RESULTS AND DISCUSSION

The Observation Method

The observation method furnishes two kinds of data - the observers' estimates of the degree of listening and the observees' actual behavior.

Degree of Listening.

For any particular program, the method provides a minute-by-minute estimate of the degree of listening of each observee. An overall measure of the degree to which a program has been listened to can be obtained by summing the number of minutes spent listening attentively, listening casually, and not listening, and then converting these figures into percentages of the total number of minutes of the program. Such a measure may be useful in comparing different programs or in comparing the same program over a period of weeks. Table II shows the percentage of time spent in the three degrees of listening for each of three programs and a miscellaneous group of programs.

TABLE II

Percentages of Time Spent in the
Three Degrees of Listening by Programs

Program	Percentage of Time Spent		
	List Att.	List Cas.	Not List
Jack Benny	33	24	43
Amos n' Andy	31	27	42
Duffy's Tavern	26	30	44
Miscellaneous Group	24	27	49

The difference between the Amos n' Andy and Jack Benny percentages in the listening attentively column is not statistically significant¹; nor is the difference between Duffy's Tavern and the Miscellaneous group. The probability that the difference between Jack Benny and Duffy's Tavern is due to chance, however, is less than .01; between Amos n' Andy or Jack Benny and the Miscellaneous group, less than .01; and between Amos n' Andy and Duffy's Tavern, less than .05.

In the listening casually column, the probability that the difference between Jack Benny and Duffy's Tavern is due to chance is less than .01. The probability that the other differences (except between Amos n' Andy and the Miscellaneous group) are due to chance is less than .05.

In the not listening column, the probability that the difference between the Miscellaneous group and Duffy's Tavern is due to chance is less than .05; between the Miscellaneous group and the other two, less than .01.

Among other things, these results clearly indicate that people do not listen to the radio all of the time the radio is on. When considered in the light of the A. C. Neilson Audimeter technique, this finding takes on added significance for it points unmistakably to the conclusion that the basic assumption underlying Audimeter ratings may not be as valid as Neilson supposes.

¹Throughout this study, we have adhered to the convention of considering a probability of .01 as highly significant, a probability of .05 as significant, and a probability of greater than .05 as not significant.

The Audimeter measures radio behavior rather than audience behavior, and the observation method has demonstrated that the two are quite different. Furthermore, the possibility that a general corrective factor might be applied to Audimeter ratings seems doubtful in view of the significant program to program differences reported above. It may be possible, however, to find specific corrective factors for each program, but the answer is to be found only by more research.

Program Profiles.

An effective way of presenting minute-by-minute degree of listening involves the use of program profiles, a technique used by investigators connected with the Lazarsfeld-Stanton Program Analyzer (11). Profiles constructed for the Jack Benny and Amos n' Andy programs show for each minute of the programs how many people were listening attentively, (listening casually,) and how many were not listening.

Figure 1 is the profile of the Jack Benny program based upon 54 observees. The upper graph indicates the number of people who were listening attentively during each minute and the lower graph indicates the number who were not listening during each minute. The number who were listening casually is not actually shown on the graph but it can be readily determined from the graph¹. Below the graphs is a minute-by-minute outline of what came over the air during the program.

¹Sum the number of people who were listening attentively and the number who were not listening during any particular minute of the program. Subtract this sum from the total number of observees and the difference is the number who were listening casually during that minute. For example, in Figure 1 at 7:10, 15 people were listening attentively and 21 were not listening, a total of 36. Subtract 36 from 54, the total number of observees. The difference, 18 were listening casually during that minute.

Commercial
 Introd; Dialogue-Rochester
 Dialogue-Rochester, Jack
 Dialogue-Rochester, Jack
 Dialogue-Rochester, Jack
 Dialogue-Mary and Jack
 Dialogue-Mary and Jack
 Dialogue-Mary and Jack
 Dialogue-Mary, Jack, Dennis
 Dialogue-Mary, Jack, Dennis
 Dennis sings
 Dennis sings; Dialogue
 Dialogue-Jack, Mary, Don
 Dialogue about quartet
 Dialogue-Jack, Mary, Phil
 Dialogue-Jack, Mary, Phil
 Dialogue-Jack, Mary, Agent
 Dialogue-Jack, Mary, Agent
 A quartet auditions
 Dialogue-Jack, Mary, Agent
 Dialogue-Jack, Mary, Agent
 Dialogue; Girl quartet
 Dialogue-Jack, Mary, Agent
 Dialogue-Jack, Mary, Agent
 Dialogue, Russian quartet
 Russian quartet singe
 Dialogue; Jack, Mary exit
 Music; 4H Club Announce
 Commercial
 Dialogue-Don, Mary; Signoff

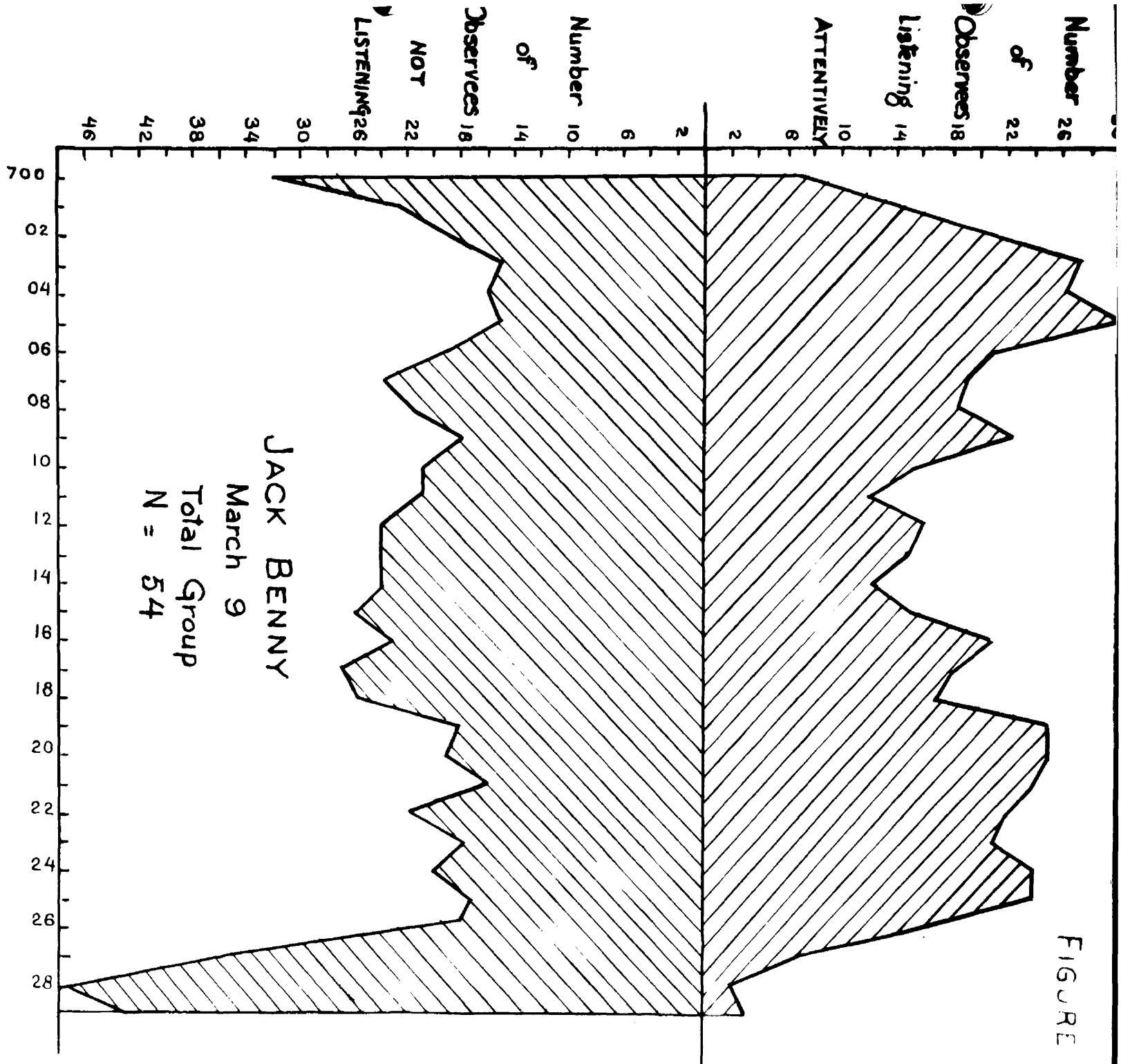


FIGURE 1

Music, Commercial, Introd.
 Dialogue-Kingfish and wife
 Dialogue-Kingfish and wife
 Dialogue-Kingfish, Insur.Agent
 Dialogue-Kingfish, Insur.Agent
 Dialogue-Kingfish and Andy
 Dialogue-Kingfish and Andy
 Dialogue-Kingfish and Andy
 Dialogue; Mills Bros. sing
 Mills Bros. sing
 Mills Bros. sing; Commercial
 Commercial; Dialogue
 Dialogue-Kingfish and wife
 Dialogue-Kingfish and Andy
 Dialogue-Kingfish and Andy
 Dialogue-Kingfish and Andy
 Dialogue-Kingfish and Andy
 Dialogue-Kingfish and Andy
 Dialogue-Kingfish and Andy
 Dialogue-Amos and Andy
 Commercial
 Dialogue-Kingfish and Agent
 Dialogue-Kingfish and Agent
 Dialogue-Kingfish and lawyer
 Dialogue-Kingfish and lawyer
 Dialogue-Kingfish and Andy
 Dialogue-Kingfish and Andy
 Dialogue-Kingfish, Andy, Doctor
 Dialogue; Commercial
 Commercial, Goodnight, Commercial
 Commercial; Music; Signoff

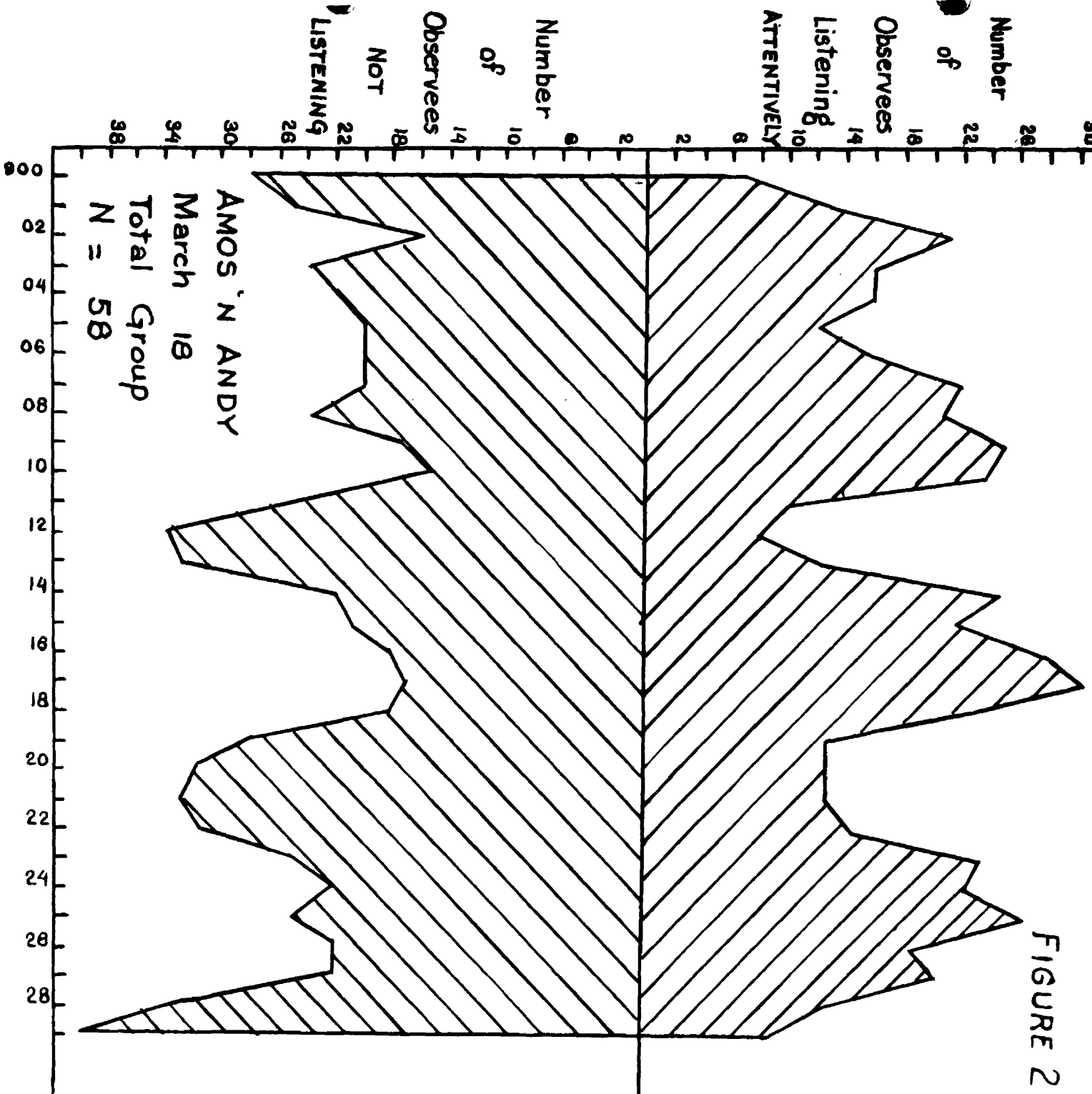


FIGURE 2

By comparing this abbreviated script with the ups and downs in the graph, one may get some idea of which parts of the program the audience attended to.

Referring to the Jack Benny profile, notice that there were two peak periods of attentive listening - one during the early part of the program and one during the late part. Looking at the profile in somewhat more detail, we see a gradual rise in attention at the beginning of the program - during the Rochester dialogue - to a peak at 7:04. From that point, attention dropped and except for a momentary rise during the Dennis Day dialogue, continued downward even while Dennis Day sang. The appearance of Phil Harris at 7:14 was followed by a slight negative effect but the discussion about the quartet at 7:16 was accompanied by an increase in attention. Except for a drop between 7:17 and 7:18 while a quartet was singing, attention was definitely on the upswing. In spite of several brief commercials in the form of verse sung by various quartets, the period from 7:19 to 7:26 was one of high, prolonged attention. The exit of Jack Benny and Mary Livingstone at 7:26, followed by music, a 4-H Club announcement, and a full commercial was accompanied by a sharp drop to a new low for the program.

Figure 2 is the profile of the Amos n' Andy program. Notice that attention built up at the beginning of the program and reached a peak at 9:02. Attention then began to wane and reached a low at 9:05 but started upward again and reached a new high at 9:09. A one minute commercial began at 9:10 and attention tumbled sharply reaching a low point at 9:12. It was not until 9:14 that attention recovered its pre-commercial level - it was

three minutes after the commercial ended before attention attained its previous high level. From there attention gained until 9:17, the high point of the program. This was followed by a drop which reached its low point from 9:19 to 9:21 - a one minute commercial had begun at 9:19 and again it was several minutes after the commercial ended before attention recovered. The last peak was reached at 9:25 and from then on attention declined. Its decline was hastened by the last commercial which began at 9:27½. Notice that the major peaks in the listening attentively graph are accompanied by drops in the not listening graph. Similarly, depressions in the listening attentively graph are accompanied by boosts in the not listening graph.

This profile, in addition to pointing up those portions of the program which were not well attended to, also suggests an interesting hypothesis about commercials, namely, that they exert a negative influence on attention which extends for several minutes after the commercial itself ends. An analysis of the observees actual listening behavior suggests that this hypothesis may be further refined, i.e., that the degree of negative influence is a function, in part, of the length of the commercial. You will recall that the several short commercials in the Jack Benny script between 7:19 and 7:26 were not followed by drops in attention. This was probably due to the way the commercials were presented as well as to the brevity of any one of them. One should check experimentally the relative importance of length of commercial and manner of presentation. Comparison of the Jack Benny and Amos n' Andy profiles suggests, however, that commercials of the kind embodied in the Jack Benny program are more

conducive to the maintenance of attention than are the full length commercials found in the Amos n' Andy program.

Reliability of Degree of Listening.

The reliability of the degree of listening has been estimated in five different ways. The first four are of the split-half variety. The Amos n' Andy audience sample, the Jack Benny sample, and the Miscellaneous sample were each divided into two groups roughly equated according to age, sex, and listening environment (home or school) of the observees. The Miscellaneous subgroups were equated exactly on the basis of program and insofar as possible on the basis of the other variables.

1. Table III shows the total amount of time spent listening attentively, listening casually, and not listening by program subgroups.

TABLE III

Percentage of Time Spent in the Three Degrees
of Listening by Program Subgroups

<u>Programs</u>	<u>Number of Observees</u>	<u>Percentage of Time spent</u>		
		<u>List Att.</u>	<u>List Cas.</u>	<u>Not List</u>
Jack Benny Group 1	27	34	24	42
Jack Benny Group 2	27	32	24	44
Amos n' Andy Group 1	29	31	27	42
Amos n' Andy Group 2	29	31	26	43
Miscellaneous Group 1	35	26	27	47
Miscellaneous Group 2	35	22	26	52
All Group 1's	91	30	26	44
All Group 2's	91	29	26	45

FIGURE 3

Number of Observers Listening ATTENTIVELY

16
14
12
10
8
6
4
2
2
4
6
8
10
12
14
16
18
9:00 02 04 06 08 10 12 14 16 18 20 22 24 26 28

AMOS
'N
ANDY
March 18

Number of Observers NOT LISTENING

Group 1
N = 29

Number of Observers Listening ATTENTIVELY

14
12
10
8
6
4
2
2
4
6
8
10
12
14
16
18
20
9:00 02 04 06 08 10 12 14 16 18 20 22 24 26 28

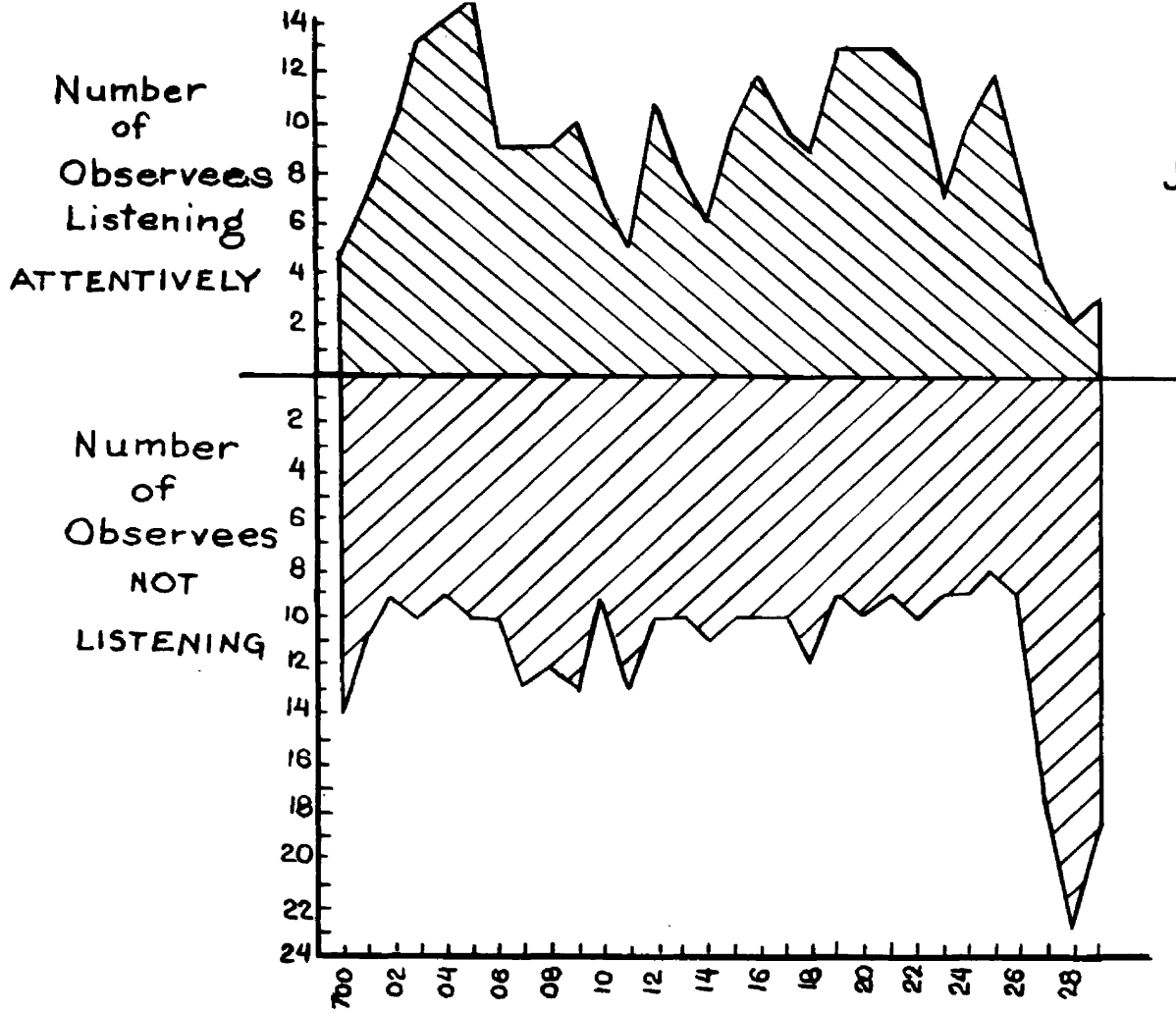
AMOS
'N
ANDY
March 18

Number of Observers NOT LISTENING

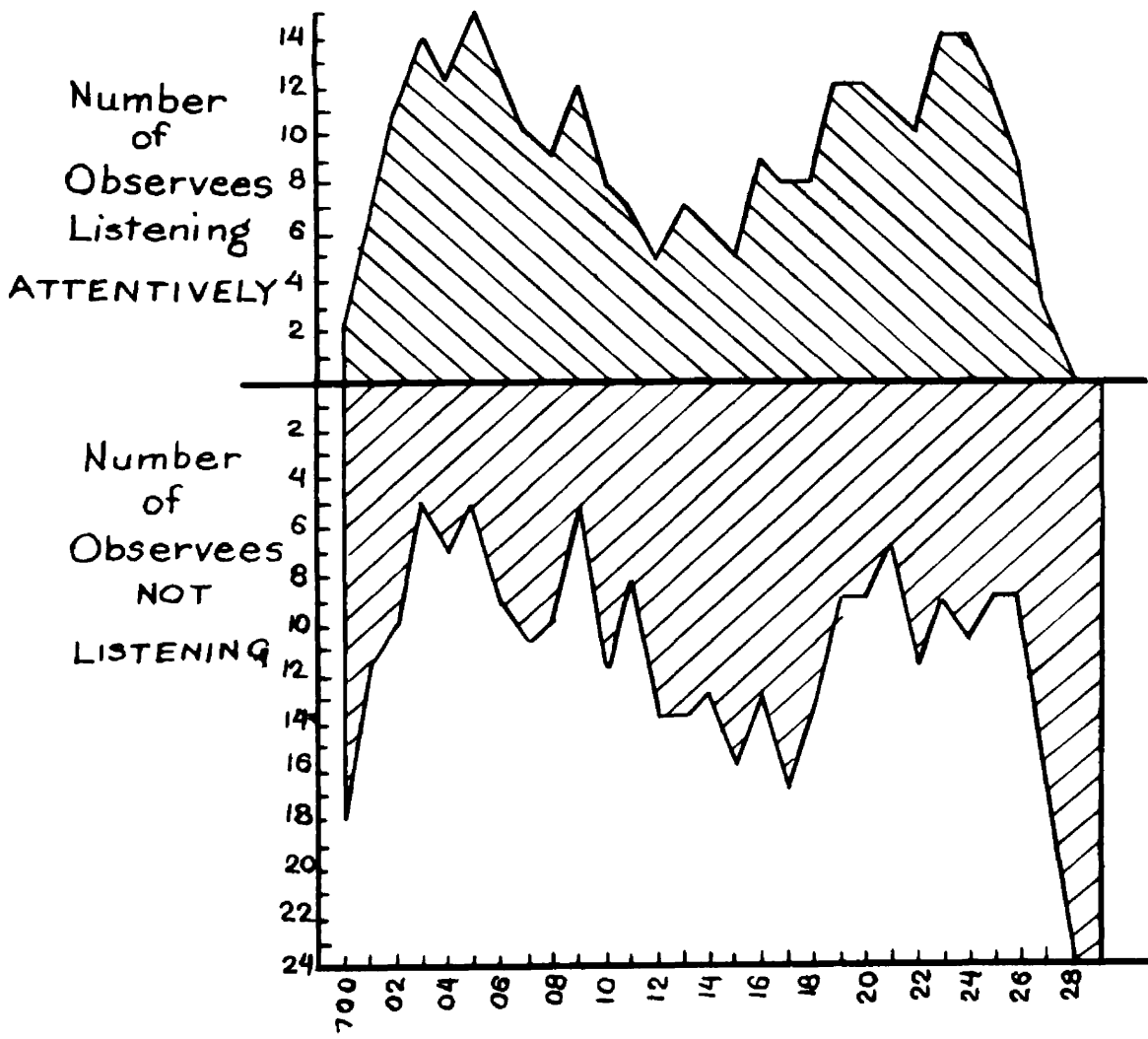
Group 2
N = 29

FIGURE 4

JACK BENNY
March 9
Group - 1
N = 27



JACK BENNY
March 9
Group - 2
N = 27



Except for the not-too-well equated Miscellaneous subgroups, the agreement is remarkably close.

2. A second estimate of reliability was based upon a listening score obtained by assigning a value of plus one to listening attentively, zero to listening casually, and minus one to not listening. Listening scores for each minute of a program were calculated for each subgroup, a total of thirty pairs of scores for a thirty minute program. Correlation of these pairs of scores yielded a value of .69 for the Amos n' Andy subgroups, and .78 for the Jack Benny subgroups.

3. A third estimate of reliability was based upon the minute-by-minute agreement between the number of observees listening attentively, listening casually, and not listening in each subgroup. For example, during the first minute of the Amos n' Andy program, 10.35% of group 1 and 13.80% of group 2 were listening attentively - a difference of 3.45%. Similarly, a percent difference was calculated for each of the other 29 minutes of the program, and then, the average percent difference was calculated. This average percent difference may be thought of as being the average percent of disagreement and the difference between it and 1.00 may be thought of as being the average percent of agreement or overlap. The average percent of agreement was determined for the three degrees of listening in both the Amos n' Andy and Jack Benny subgroups. They are shown in Table IV.

TABLE IV

Coefficients of Determination
(Percents of Agreement)

<u>Programs</u>	<u>List Att.</u>	<u>List Cas.</u>	<u>Not List.</u>
Amos n' Andy	.94	.92	.90
Jack Benny	.92	.90	.90

4. A fourth estimate of reliability is graphic and merely involves superimposing and comparing two profiles of the same program - subgroup profiles. Figures 3 and 4 are the profiles of the two Amos n' Andy and the two Jack Benny groups, respectively. Although the profiles of the subgroups differ in detail, their basic patterns are unmistakably similar.

5. The last estimate of reliability was based upon the double observer data. Six pairs of siblings observed the radio listening behavior of 19 people for a total of 736 observee minutes; each observee minute provided two independent estimates of degree of listening. Of these 736 paired estimates, there was perfect agreement in 571 or 69% of the cases. In 198 or 27% of the cases, the estimates were off one category - one member of a pair estimated listening casually and the other estimated listening attentively or not listening. In 27 or 4% of the cases, the estimates were off two categories - one member of a pair estimated listening attentively and the other estimated not listening. As stated above, the percentage of complete agreement was 69%. The square root of 69% is .83, the estimated coefficient of reliability.

In summary, three tests yielded reliability coefficients ranging from .69 to .83. A fourth test revealed remarkably close

agreement between matched groups of listeners in the percentage of time spent in the three degrees of listening. A fifth test indicated a definite similarity between the program profiles of equated groups of listeners.

Observation Method - Hooper Telephone Interview Cross Check.

Hooper interviewers obtain two basic facts from respondents. First, they ask if the respondent was listening, and second, they ask how many people in the house were listening. These two facts serve as the basis for the widely circulated network Hooperatings and the audience composition reports produced by the Hooper organization. A cross check between the observation method and the Hooper method enables us to compare the two basic facts as reported over the telephone and as reported in writing by on-the-scene observers.

Was the Respondent Listening?

All twenty¹ respondents reported over the telephone that they were listening to the radio at the time of the call. According to the observers, eight² of the twenty were listening attentively, one was listening casually, and seven were not listening. In the other four cases, the observation data did not reveal who had answered the phone - these four may not have been

¹Actually 60 homes were called and 60 students had agreed to observe at that time. The following day, however, it was learned that only twenty students had been able to observe and thus, the number of cross comparisons possible was reduced to twenty.

²This includes four observers who happened to answer the telephone and who were, therefore, respondents also.

part of the observer groups because of their failure ever to enter the rooms with the radios.

Of the seven who were judged as not listening, four were out of the room, two were reading, and one was conversing about a subject unrelated to the radio program. Seven out of sixteen, almost half, reported that they were listening to the radio but were reported by the observers as not listening at the time of the call.

How many persons in the house were listening?

The twenty respondents reported a total of sixty persons listening to the radio at the time of the call. According to our observers, 27% of these were listening attentively, 16% were listening casually, and 17% were not listening. In terms of percentages, 45% were listening attentively, 27% were listening casually, and 28% were not listening.

This is what the 17 non-listeners were doing: four were out of the room, 3 were asleep, 3 were conversing about subjects unrelated to the program, 2 were reading, and 4 were in the room with the radio but the radio had been turned off eleven minutes before the telephone rang.

One telephone respondent reported that no one was listening to the radio, but the observer in that home reported two people (including himself) listening attentively.

This includes all twenty observers whom it was assumed were listening attentively.

If listening attentively and listening casually are lumped together and treated as listening, then in ten of the twenty homes there was perfect agreement between what was reported over the telephone and what was reported by the observers. In each of the other ten homes, the respondent reported at least one person listening who was reported as not listening by the observer. Of the four cases in which the telephone respondent and the observer was the same person, one failed to be consistent in his telephone and recorded reports.

The Observation method-Heoper method cross check has indicated, then, that reports from a telephone respondent and reports from an observer do not even approximate complete agreement. Should these results be substantiated on larger samples, they will literally strike the coincidental out of the Heoper coincidental interview, for such an interview may no longer be thought of as accurately reflecting behavior at the time of the call. The following passage was taken from a discussion by Chappell and Heoper of the validity of the coincidental interview (2).

The first requirement of any method for measuring radio audience size in any selected population is that it will accurately reflect what people do. The coincidental method asks, "Were you listening to the radio just now?" No memory is involved. The question is asked of the person who knows the answer, and he reports directly concerning listening behavior. This is particularly important, as will become apparent when other methods are considered. In the case of the mechanical recorder,..... for example, behavior is measured directly. But the behavior measured is set-tuning not listening. Listening behavior is inferred in the recorder methods. No inference is involved in the coincidental method. There can be no appeal from the respondents report. It is, therefore, difficult to conceive of a method that could measure listening behavior with greater validity.

The Observation data suggest, of course, that the Hooper method grossly overestimates the size of the listening audience. If further investigation should reveal that this overestimate varies irregularly from program to program, then the present Hooper ratings would be an even greater misrepresentation of the actual facts because they would not even reflect relative audiences. Such a finding would also preclude the possibility of using a general corrective factor with the Hooper ratings. As suggested in the discussion of the Heilson ratings, the possibility of finding corrective factors for specific programs can only be confirmed or negated by more research.

Analysis of Observers Actual Behavior.

In addition to the observers' estimate of degree of listening, the observation method provides another kind of data - the actual behavior of the observees, as recorded by the observers. The observations in this study were collected under two environmental situations, in the home and at school, but because of its greater import the analysis have been confined almost exclusively to the data gathered in the home.

Seven classes of activities occurred with sufficient frequency to warrant a detailed tabulation.

1. Conversing about subject unrelated to the program.
2. Reading or writing.
3. Being out of the room.
4. Sleeping in the room with the radio.
5. Sewing, knitting, or crocheting.
6. Positive program behavior - smiling, laughing, grinning, singing or humming a tune coming over the air, conversing about the program either favorably or neutrally, or engaging in any other behavior which is directly related to the radio program and which is not antagonistic to it.

7. Negative program behavior - switching the dial, turning off the radio, commenting unfavorably about the program, or engaging in any other behavior which is directly related to the radio program and which expresses dislike of some phase or all of it.

The number of minutes during which any of these activities occurred was tabulated. If an observee was reading during any part of a minute, that minute was tabulated as a reading minute. If more than one activity was going on during the same minute, that minute was tabulated in more than one place. Tabulations were made by program, sex, degree of listening, and listening environment.

Table V-A shows the percentage of observed minutes of each program during which each of these activities occurred. These figures refer to the home listening situation only. Table V-B shows the probabilities that the differences found in Table V-A are due to chance.

TABLE V-A

Percentage of Observed Minutes of Each Program
During Which Each Activity Occurred

Program	Con- vers- ing	Reading Writing	Out of Room	Sleep- ing	Sev- ing	Positive Behavior	Negative Behavior	Total
Amos n' Andy	16.8	13.0	8.6	1.5	3.9	26.2	0.5	70.5
Jack Benny	10.3	11.8	7.7	2.4	0.0	27.1	1.9	61.2
Duffy's Tavern	8.9	16.3	15.1	1.4	8.1	21.1	0.5	71.4
Miscell. Group	13.5	16.7	6.8	1.0	1.9	10.3	1.1	51.3

TABLE V-B

Probabilities of the Significance of the Differences in the Percentage of Observed Minutes During Which Each Activity Occurred

Differences Between	Con- vers- ing	Reading- Writing	Out of Room	Sleep- ing	Sew- ing	Positive Behavior	Negative Behavior
Amos n' Andy and Jack Benny	.01	-*	-	.05	.01	-	.01
Amos n' Andy and Duffy's Tavern	.01	-	.01	-	.01	.05	-
Amos n' Andy and Miscellaneous	.05	.01	-	-	.01	.01	-
Jack Benny and Duffy's Tavern	-	.05	.01	-	.01	.05	.05
Jack Benny and Miscellaneous	.05	.01	-	.05	.01	.01	-
Duffy's Tavern and Miscellaneous	.01	-	.01	-	.01	.01	-

The extreme right hand column of Table V-A shows the number of observed minutes accounted for by the seven activities. Because of some overlap, these figures are actually a slight over-estimate of the amount of time accounted for. Some of the unaccounted for time was due to the failure of every observer to record the actual behavior of every observee every minute. Many judgmental rather than behavioral entries appear in the data - entries such as, "Dad is listening," "Mother doesn't like the program," "Ann isn't listening," "John likes the program."

The bulk of the residual time was taken up in the wide variety

* Greater than .05.

of activities in which people engage while presumably listening to the radio. While reading the observation records, one cannot help but be impressed by the tremendous variability from individual to individual in the amount of time spent in different activities. Here is a partial list of the things our observers noted people doing: dancing, playing cards, staring into space, playing with the cat or dog, washing dishes, combing ones hair, ironing clothes, eating, cleaning the room, playing the piano or organ, serving candy, washing ones hair, cleaning ones finger nails, polishing ones finger nails, cutting ones toe nails, coughing and sneezing, singing, undressing, smoking, fooling around with a coconut, fixing the clock, scratching ones back, cooking, and so on.

Relationship between actual Behavior and Estimate Degree of Listening.

In order to relate the observers estimates of degree of listening to the actual behavior of the observees, the seven activities were tabulated by degree of listening. Table VI shows the percentage of each activity which occurred during each degree of listening.

TABLE VI

Percentage of Each Activity Occurring During Attentive Listening, Casual Listening, and Non-Listening

Activity	Listening Attentively	Listening Casually	Not Listening
Conversing	2	21	77
Reading or Writing	2	46	52
Out of Room	0	0	100
Sleeping	0	0	100
Sewing	24	46	28
Positive Behavior	57	13	0
Negative Behavior	60	18	22

Seventy-seven percent of the minutes during which observees conversed about subjects unrelated to the program, and 92% of the minutes during which they read or wrote were categorized as "not listening" minutes. All the minutes spent out of the room or asleep were so categorized.

The reader may wonder that 22% of the negative behavior occurred during minutes which were classified by the observers as "not listening" minutes. In many instances the observees did such things as turning the radio off, turning the volume way down, or telling someone else to turn the radio off, and although they may have been aware that the radio was on, the observers felt that they were not attending to it.

The reader may also be disturbed by the fact that 13% of the positive behavior occurred during minutes which were categorized as "listening casually" rather than "listening attentively" minutes. There were many minutes during which observees sang, whistled, hummed, or tapped their feet in time with music coming over the air, but the observers were not sure that the observees were necessarily listening attentively and so they classified them as listening casually.

Analysis of Behavior by Sex of Observees.

Table VII shows the breakdown of the various activities by the sex of the observees. For each activity, Table VII indicates the percentage of male-observee minutes and the percentage of female-observee minutes during which the activity occurred. The probability of any of the sex differences being due to chance is also shown.

TABLE VII

Percentage of Male-Observed Minutes and Female-Observed Minutes During which Each Activity Occurred and the Probability That the Sex Differences are Due to Chance

Program	Con- versing	Reading Writing	Out of Room	Sleeping	Positive Behavior	Negative Behavior
Amos n' Andy						
% Male	10.7	19.5	2.2	*	29.8	*
% Female	18.7	6.8	13.6	*	23.5	*
p diff.	.01	.01	.01		.05	
Jack Benny						
% Male	9.2	14.2	8.0	*	27.7	*
% Female	11.3	9.9	7.5	*	26.4	*
p diff.	-	.03	-		-	*
Duffy's Tavern						
% Male	4.8	16.7	18.6	*	25.7	*
% Female	11.4	16.1	13.1	*	18.3	*
p diff.	.01	-	-		.05	
Miscellaneous						
% Male	12.3	20.5	5.1	*	9.0	*
% Female	14.6	13.7	8.2	*	11.5	*
p diff.	-	.01	.05		-	
Total						
% Male	10.4	18.2	6.4	1.9	21.1	1.1
% Female	15.1	11.3	10.0	1.3	19.4	1.1
p diff.	.01	.01	.01	-	-	-

* too few cases to justify a breakdown

- not significant

Examination of Table VII suggests some interesting differences between the sexes. The data suggest that women spend more time conversing about things unrelated to the programs than do men. Although the differences for the Jack Benny program and for the Miscellaneous group are not significant, they are in the same direction as are the highly significant differences for the Amos n' Andy and Duffy's Tavern programs. For all the

Programs combined, the difference is highly significant.

The reverse of this appears in the case of reading and writing. Men read and write more than do the women. All the differences are significant except for the Duffy's Tavern program and it is in the same direction as the others. The probability that the difference for all programs combined is due to chance is less than one in a hundred.

Although for the total group, women spent significantly more time out of the room than did men, there are enough inconsistencies from program to program to forestall any generalization about this aspect of listening behavior.

No other highly significant, or consistent differences appear in Table VII.

Relationship between Commercials and Conversation.

To test the hypothesis that more conversation takes place during and immediately after commercial announcements than during other parts of a program, the number of observers conversing during each minute of the Ames a' Andy program was tabulated.

Four commercial announcements were made during this particular program. The first one was made during the first minute of the program and lasted about 15 seconds. The other three were made at about 9:10, 9:19, and 9:27 respectively and each lasted a minute.

The thirty minutes of the program were divided into two groups, a "commercial minutes" group and a "non-commercial minutes" group. The former was made up of thirteen minutes which included the four minutes during which the commercials were read,

the one minute following the first, short commercial, the three minute following the second and third commercials, and the two minutes - the remainder of the program - following the last commercial. The "non-commercial" group included the remaining seventeen minutes of the thirty minute program. Table VIII presents the distributions of the number of observees conversing varying numbers of minutes for each of these two groups. The mean and standard deviation for each group is also shown.

TABLE VIII

Distributions of the Number of Observees Conversing Varying Numbers of Minutes for the "Commercial-Minutes" Group and the "Non-Commercial-Minutes" Group

Number of Observees Conversing	"Commercial- Minutes"	"Non-Commercial- Minutes"
13 or more minutes	2	0
11 - 12	4	0
9 - 10	5	1
7 - 8	0	5
5 - 6	2	6
3 - 4	0	3
1 - 2	0	2
	13	17
	10.6	5.6
	3.3	2.2
		Mean
		Standard Devia-
		tion.

Notice how little these two distributions overlap. The probability that the difference between the two means occurred through random errors of sampling is less than one in a hundred, Student's *t* being equal to 5.0. This bears out the hypothesis that more conversation takes place during and immediately after commercials than during other parts of this Amos n' Andy program. It was not possible to test this hypothesis with the Jack Benny program or with the Duffy's Tavern program. The Jack Benny

Program had no extended commercials during the body of the program, and the total number of observees for the Duffy's Tavern program was too small to enable a minute-by-minute breakdown of the frequency of conversation.

The Coincidental Telephone Interview Method

The second major phase of this research was concerned with the adaptation of the coincidental telephone interview method to the determination of audience reactions to radio programs.

Sixty eight interviewers called a total of 1020 homes of families of University of Maryland students. Table IX shows the outcome of these calls. Column 1 indicates the number of homes in which respondents were listening to various programs, the number in which respondents were listening to the radio but were unable to identify the program being listened to, the number in which respondents were not listening to the radio, the number in which no one answered the telephone, and the number in which a busy signal was obtained. In accordance with Hooper policy, the busy numbers were distributed among listening and not listening in the same proportions as were the rest of the sample. The figures obtained after this distribution of busy numbers is shown in column 2. The last column of Table IX shows the percentage of all homes called in which people reported that they were listening to various programs, the percentage in which people reported that they were not listening to the radio, and the percentage in which no one answered the telephone. These percentages are based on 1017 homes - homes in which telephones have been disconnected are discarded from Hooper samples.

The percentages shown for the Amos n' Andy and Vox Pop programs are the Hooperatings of those programs as determined from our sample.

TABLE IX

Outcome of 1020 Coincidental Interviews

	Number of Homes	Number of Homes After Apportion- ment of Busy Lines	Percentage * of all Homes Called
Listening to Amos n' Andy	254	276	27.1
Listening to Vox Pop	52	56	5.5
Listening to other programs	86	93	9.1
Listening but didn't know program	33	36	3.5
Not listening to radio	412	447	44.0
No answer	109	109	10.7
Busy lines	71	-	-
Disconnected tele- phones	3	3	-
	1020	1020	100.0

* Based on N of 1017 - the 3 disconnected telephones were discarded from the sample.

It is of interest to compare the Hooperatings of the Amos n' Andy and Vox Pop programs obtained in this study with the national ratings obtained for these programs by the Hooper organization. Although Hooper did not check these programs on March 25, 1947, he did check them the week before and the week after. The average of these two Hooper checks provide an esti-

state of what the ratings might have been had they been checked on March 25, 1947. Table X shows these averaged national Hooperatings and the ratings obtained in this study.

TABLE X

Comparison of National Hooperatings with
Hooperatings Obtained in this Study

Program	National Hooperatings Averages of March 18 and April 1	Hooperatings Obtained in this Study March 25
Amos n' Andy	23.9	27.1
Vox Pop	7.6	5.5

The national Hooperating for the Amos n' Andy program was 3.2 less than the Amos n' Andy rating as determined in this study. The probability of obtaining a difference that large falls well within the expectancy of chance. On the other hand, the critical ratio of the difference between the Vox Pop ratings is just significant at the .05 level. The differences in the two ratings, however, are small enough to suggest that the samples used throughout this research - in both the observation and telephone phases - may not be too unrepresentative of what one might find throughout the nation.

Analysis of Answers to Question 3 of the Coincidental Interview.

Question 3 of the coincidental interview asked respondents whether "compared with other Amos n' Andy (or Vox Pop, etc) programs, tonight's program so far is excellent, good, or just fair?" Answers to this question have been analyzed for the 254 respondents who were listening to Amos n' Andy and for the 52

respondents who were listening to Vox Pop. Table XI shows how the listeners to these programs answered the question.

TABLE XI

Responses to Question 3 of the
Coincidental Interview

Response	Amos n' Andy Respondents		Vox Pop Respondents	
	N	%	N	%
Excellent	61	24.0	20	38.5
Good	111	43.7	19	36.5
Just Fair	42	16.5	7	13.5
Was not listening	18	7.1	4	7.7
Average	8	3.1	0	0.0
Don't know	9	3.5	1	1.9
No answer	5	2.0	1	1.9
	<hr/> 254	<hr/> 100.0	<hr/> 52	<hr/> 100.0

The figures in Table XI reveal an interesting fact. The listeners to Vox Pop - a low cooperating program - seem to think more of the Vox Pop program than the listeners to Amos n' Andy - a high cooperating program - think of Amos n' Andy. Twenty four percent of the Amos n' Andy listeners said the program was excellent; 38.5% of the Vox Pop listeners said the Vox Pop program was excellent. The probability that this difference is due to chance is less than .05.

Notice that 7.1% of the Amos n' Andy respondents and 7.7% of the Vox Pop respondents said that they were not listening - after having said that they were listening in answer to questions 1 and 2 of the interview. Some of these people probably were not

listening very attentively but others may have given that answer to avoid having to make a judgment about the program.

Table XII shows the way men and women respondents categorized the Amos n' Andy program in answer to Question 3. The Table contains only those respondents who answered excellent, good, or just fair.

TABLE XII

Answers to Question 3 by Sex of
Amos n' Andy Respondents

	Men		Women	
	N	%	N	%
Excellent	33	38.8	26	22.2
Good	39	45.9	65	55.6
Just Fair	13	15.3	26	22.2
	<hr/>	<hr/>	<hr/>	<hr/>
	85	100.0	117	100.0

Chi Square for this table is 6.80. The probability of obtaining a value that high or higher through random errors of sampling is less than .05 and greater than .02, thereby supporting the statement that women were less favorably inclined toward the program than were the men.

To test the hypothesis that the frequencies of the responses to Question 3 vary significantly during different portions of the program, the responses for the Amos n' Andy program were tabulated according to the times they were made. The period from 9:07 to 9:30 was divided into five groups: 9:07 to 9:10, 9:11 to 9:15, 9:16 to 9:20, 9:21 to 9:25, and 9:26 to 9:30. Table XIII shows the frequency of each response during each of these five periods.

TABLE XIII

Frequency of Amos n' Andy Listeners' Responses
to Question 3 by Time Periods

	Time Period				
	<u>9:07-9:10</u>	<u>9:11-9:15</u>	<u>9:16-9:20</u>	<u>9:21-9:25</u>	<u>9:26-9:30</u>
Excellent	9	17	15	12	8
Good	23	21	19	28	20
Just Fair	7	10	10	11	4
Other	5	9	12	10	4
Totals	44	57	56	61	36

Chi Square for this table was calculated and it fell well within the range of chance expectancy, thus negating the hypothesis that frequencies varied significantly during different parts of the program.

Reliability of Answers to Question 3 by Amos n' Andy Respondents.

The 254 interview sheets were arranged in order of the times the interviews were made. The sheets were then divided into two piles by putting the first sheet in pile 1, the second in pile 2, the third in pile 1, the fourth in pile 2, and so on. The result was two groups of 127 interviews each. Table XIV shows the frequency of responses to Question 3 for each of these two groups.

TABLE XIV

Frequency of Responses to Question 3
by Amos n' Andy Subgroups

	Excellent	Good	Just Fair	Other	Total
Amos n' Andy Group 1	30	54	20	23	127
Amos n' Andy Group 2	31	57	22	17	127
Total	61	111	42	40	254

None of the differences in this Table is significant, thus indicating the reliability of the overall responses.

Analysis of Question 4 and 5 of the Coincidental Interview.

Questions 4 and 5 were intended to induce respondents to state what they believed to be the worst and best parts of the program to which they were listening at the time of the call. Answers to these two questions were tabulated and categorized.

Table XV indicates for the Acee n' Andy listeners, the frequencies of various responses to Question 4 - what was the worst part of tonight's program?

TABLE XV

Acee n' Andy Listeners Responses
to Question 4

Categorized responses	Number	%
There is no worst part	78	31
No answer	48	19
Don't know	33	13
Not listening attentively enough to answer	39	15
Haven't listened long enough to answer	10	4
Commercials	17	7
Trailer incident	5	2
Don't like any of it	3	1
Other answers	21	8
	254	100

The above Table contains no wealth of particularly useful information. The only "worst parts" which were mentioned with any frequency were the commercials and the trailer incident. The great bulk of respondents either couldn't or wouldn't answer the question, or they reported that there was no worst part.

Table XVI shows, for the same program, the answers to Question 5 - what was the best part of tonight's program?

TABLE XVI

Amos n' Andy Listeners Responses
to Question 5

Categorized responses	Number	%
No answer	50	20
It was all good	42	17
Not listening attentively enough to answer	28	11
Haven't listened long enough to answer	9	4
Don't know	16	6
No outstanding part	11	4
Trailer incident	28	11
The Kingfish	12	5
The singing quartet	15	6
The jokes, comedy, humor	10	4
The lawyer and his eye-glasses	2	1
Amos making a fool out of Andy	1	1
Amos	1	1
The negro dialect	2	1
The characters	2	1
Other answers	17	7
	<hr/> 254	<hr/> 100

As can be seen from Table XVI, the respondents were more informative in answering Question 5 than they were in answering Question 4. Notice the drop in frequency of the "not listening attentively enough to answer," and the "don't know" categories, as compared with comparable figures in Table XV. The same respondents who were not listening attentively enough to answer Question 4 were listening attentively enough to answer Question 5. The trailer incident was the most frequently mentioned "best part." Other specific items with more than one mention were the Kingfish, the singing quartet, the jokes, the lawyer, Amos, the negro dialect, and the characters. The bulk of the "other answers" were specific but because they occurred only once each, they are not indicated in the Table. In all, however, almost 40% of the respondents gave a reasonably specific answer to the

question.

Similar results were encountered in the analysis of answers of the Vox Pop listeners. Table XVII gives the responses to Question 4.

TABLE XVII
Vox Pop Listeners Responses
to Question 4

Categorized responses	Number	%
There is no worst part	20	38
Don't know	8	15
Not listening attentively enough to answer	6	12
No answer	6	12
Haven't listened long enough to answer	2	4
Commercials	3	6
Master of Ceremonies	1	2
Other answers	6	12
	<hr/> 52	<hr/> 100

As with the Amos n' Andy program listeners, the most frequently mentioned specific item was the commercial. Six percent of the Vox Pop listeners mentioned it as compared with seven percent of the Amos n' Andy listeners. That only six percent of the vox pop listeners mentioned it takes on added significance at this time because of recent dispute between the sponsor and the producers of the program over the number of commercials being read. The producers felt that there were too many; the sponsor said no. The result was the severance of a contract. Yet the Vox Pop listeners in this small study didn't seem to be particularly bothered by the commercials - only three were bothered enough to mention it.

Table XVIII shows the responses of the Vox Pop listeners to Question 5.

TABLE XVIII

Vox Pop Listeners Responses
to Question 5

Categorized responses	Number	%
It was all good	11	21
Not listening attentively enough to answer	6	12
No answer	4	8
Don't know	4	8
There was no best part	2	4
The teacher	9	17
The part about the Museum of Natural History	2	4
The Indian boy	2	4
Other answers	12	23
	52	100

The teacher, one of the contestants on the program, seems to have been well regarded by the respondents. The Indian boy and the part about the museum each received two favorable votes.

In summary, the results indicate that question 5 - the excellent, good, just fair question - served its purpose well. This question may be of real value in differentiating the way an audience receives different programs or the same programs at different times. On the other hand, questions 4 and 5 were not nearly as valuable although question 5 did considerably better than question 4. Much of the failure of these questions can probably be attributed to the gross inexperience of the interviewers. Probing for answers may have been too much of a task for untrained interviewers. Some of the fault, no doubt, lies in the questions themselves. Also, it may be that the people who listen to a program think well of it and, therefore, would not be likely to be too critical of it. In conclusion, the coincidental telephone interview method may be of real use in providing a broad measure of what listeners think about a radio

program. The question of whether or not the method may yield useful, specific suggestions about the good and poor parts of a program will be answered only by further research.

CHAPTER IV

SUMMARY AND CONCLUSIONS

This research was designed to adapt two methods - the coincidental telephone interview method and the observation method - to the determination of how people react to radio programs.

Using the coincidental telephone interview method, sixty eight University of Maryland students conducted a five question, minute-and-a-half interview with a total of 1020 respondents. The first two questions were taken from the standard Hooper interview and the last three were designed to induce respondents to verbalize their opinions about a radio program. The Hooper ratings calculated from the data of this study agreed closely with the Hooper ratings obtained from a 33 city sample by the Hooper organization, suggesting that the samples used throughout both the telephone and observation phases of this investigation may not be too unrepresentative of a national audience.

Analysis of interview question 3 - the question which required respondents to make a judgment as to whether the program was excellent, good, or just fair - showed a satisfactory spread of responses. It was pointed out that the answers to such a question might serve to differentiate the degree to which listeners liked different programs or the same programs at different times.

Questions 4 and 5 asked respondents to name the worst and

best parts of the program, respectively. The results indicated, however, that most respondents either would not or could not verbalize their opinions, although they did have more to say about the best part of the program than about the worst part. Some of the failure to get satisfactory responses was probably due to the inexperience of the interviewers. Evidently the task of probing for answers was too much for so untrained a group. Part of the difficulty probably lies in the questions themselves. It is possible, also, that people who listen to a program are likely to be favorably inclined toward that program and, therefore, may not be the ones from whom to find out the worst parts of the program. A considerable amount of research remains to be done before questions 4 and 5 will be ready to serve their functions.

Another aspect of this research involved the development of the observation method as a measure of audience behavior. With this method, 145 University of Maryland students observed 254 people for a total of 7620 minutes of radio 'listening.' The method provided two kinds of data, namely, the observers' actual listening behavior and the observers' estimates of the degree to which the observers were listening. All of this information was provided on a minute-by-minute basis.

An overall measure of the degree to which a program was attended to was obtained by determining the total number of minutes during which the audience listened attentively, listened casually, and did not listen. Such a measure shows the relative attention holding power of a program and might be of real value in comparing different programs or in comparing the same program

at different times.

Minute-by-minute degree of listening was expressed in the form of a program profile. The profile indicates, for each minute of a program, the number of people who were listening attentively, listening casually, and not listening. By checking the ups and downs in the profile against the program script, one can determine the specific portions of a program which were accompanied by attentive or inattentive listening. Such information should be valuable to the producers or talent of a program, for it would provide them with 'what-the-audience-did' evidence of the good and poor parts of their programs.

Detailed analysis of the Aace n' Andy profile suggested the hypothesis that commercials exert a negative influence on attention which extends for several minutes after the commercial itself has ended. This hypothesis was partially substantiated by other results which demonstrated that significantly more conversation takes place during and immediately after a commercial than during other parts of a program. The data also suggested that the short commercials used in the Jack Benny program are more conducive to the maintenance of audience attention than are the full length commercials used in the body of the Aace n' Andy program.

The reliability of the estimates of degree of listening was checked in five ways. Three checks yielded reliability coefficients ranging from .69 to .83. A fourth test revealed close agreement between matched groups of listeners in the percentage of time spent listening attentively, listening casually, and not listening. A fifth estimate showed a definite similarity

between the program profiles of equated groups of listeners.

The actual behavior of the listeners, as recorded by the observers, was also analyzed. Seven categories of behavior occurred with sufficient frequency to warrant detailed tabulation. These seven were: conversing about subjects unrelated to the program, reading and writing, being out of the room, sleeping, sewing, positive program behavior, and negative program behavior. Each of these activity patterns occurred with different frequency during different programs, many of the differences between programs being statistically significant. Also, some interesting differences between the sexes in the occurrence of these activities were discovered. For example, the data indicated that women spent more time talking about things unrelated to the radio, and men spent more time reading and writing.

In a cross check between the observation method and the Hooper coincidental telephone interview method, it was found that what telephone respondents reported and what observers reported did not even approximate perfect agreement. The data suggested that the so-called coincidental interview does not accurately reflect coincidental listening and that the Hooper ratings grossly overestimate the size of the listening audience.

The observation data also pointed to the conclusion that the basic assumption underlying the A.C. Nelson Audimeter method may not be as valid as that company supposes. The Audimeter measures the behavior of the radio set rather than the behavior of the audience, and the observation method has shown that the two are quite different.

In conclusion, it has been demonstrated that the observation method can provide a satisfactorily reliable minute-by-minute record of what people do while they listen to the radio. The method furnishes a general estimate of the degree to which people listen to a program, and more specifically, it points out the particular parts of a program which were listened to attentively, listened to casually, and not listened to. Such information should be of real value to the radio industry. The observation method, furthermore, may serve another very important and useful function, namely, that of providing a criterion of audience behavior against which to validate the currently available methods of measuring audience size and audience behavior.

The initial response of the critic may be that, although the observation method has real merit, it is impractical to use on a large scale. In that connection, it can be said that the writer is currently engaged in negotiations with an organization that has a well established field staff capable of economically and expediently employing the observation method on a nationwide basis.

Future research with the observation method should follow along two avenues. First, on the basis of what is now known about the method, there appear to be certain refinements which would considerably simplify the task of the observer. For example, rather than requiring an observer to write out the more commonly occurring listener activities, it might be possible to set up a series of labeled columns so that the observer would need only to put a check mark in the appropriate place.

The second avenue of research should be concerned with the

broader problem of using observation data as the criterion against which to make large scale appraisals of the validity of the Hooper, Neilson, Program Analyzer, and Schwerin techniques. If it should be found that the Hooper and Neilson ratings do not distort the relative sizes of program audiences, then it would have been demonstrated that these methods could be depended upon to perform a most useful function. If it should be found that a program analysis and an observational analysis of the same program produce the same end result, then it would be more economical to use the cheaper Program Analyzer. The important thing is that all the methodological cross checks be made and that the capabilities and limitations of the existing methods be made known to the radio industry. The observation method can play a significant role in discovering such facts. The method of direct observation, furthermore, can supply certain kinds of information about the listening audience which no other currently available method is capable of providing.

SELECTED BIBLIOGRAPHY

1. BARTON, S. G. The radio panel. In Blankenship, A. B., How to conduct consumer and opinion research. New York: Harpers & Bros., 1946.
2. CHAPPELL, M. E., & HOOPER, C. E. Radio audience measurement. New York: Stephen Daye, 1944.
3. COLUMBIA BROADCASTING SYSTEM Does radio sell goods? New York: Columbia Broadcasting System, 1933.
4. COLUMBIA BROADCASTING SYSTEM A study of radio network popularity conducted and audited by Price, Waterhouse and Company. New York: Columbia Broadcasting System, 1930.
5. CROSSLEY, A. M. Taking the blue sky out of the air - ten years later. Advert. & Selling. July 1939.
6. CROSSLEY, A. M. Radio and sales. Variety Radio Directory. 1940, 37-93.
7. CURTIS, ALBERTA. The reliability of a report on listening habits. J. appl. Psychol., 1939, 23, 127-130.
8. CARTNELL CORPORATION Experience of 127 firms with radio broadcasting. Chicago: Cartnell Corp., 1930.
9. FEED, R. E., & CLAUSEN, J. A. The use of the mail questionnaire in social research. Paper delivered at the meetings of the American Statistical Association, Atlantic City, Jan. 24, 1947.
10. GREENE, K. Report of an NBC survey. Paper delivered at the meetings of the American Statistical Association, Atlantic City, Jan. 25, 1947.
11. HOLLOWQUIST, T., & SUSHMAN, E. A. Listening to the listeners. In Lazarsfeld, P. F., & Stanton, F., Radio Research 1942-43. New York: Duell, Sloan & Pearce, 1944.
12. HOOPER, C. E. Radio's inferiority complex. An address to Radio Executives Club of New York, Feb. 1947.
13. JOSHIMA, J. O., & GALLENBACH, E. Obliviscence during sleep and waking. Amer. J. Psychol., 1924, 35, 605-612.

14. HIRKPATRICK, C. Report of a research into the attitudes and habits of radio listeners. St. Paul: Webb Book Publishing Co., 1933.
15. LAZARSPELD, P. F. The use of mail questionnaires to ascertain the relative popularity of network stations in family listening surveys. J. appl. Psychol., 1940, 24, 802-816.
16. LAZARSPELD, P. F., & FIELD, H. The people look at radio. Chapel Hill: Univ. of North Carolina Press, 1946.
17. MEYEROWITZ, A., & FISKE, MARJORIE The relative preferences of low income groups for small stations. J. appl. Psychol., 1939, 23, 158-162.
18. NEILSON, A. C. New facts about radio research. An address to Radio Executives Club of New York, March 1946.
19. RUSLOW, S., & KELLY, N. The personal interview-roster method of radio measurement and its application. In Blankenship, A. B., How to conduct consumer & opinion research. New York: Harpers & Bros., 1946.
20. SCHERWIN, E. An exploratory study of the reliability of the "Program Analyzer." J. appl. Psychol., 1940, 24, 742-745.
21. STANTON, F. Notes on the validity of mail questionnaire returns. J. appl. Psychol., 1939, 23, 95-104.
22. SUCHMAN, E. A., & McCANDLESS, B. Who answers questionnaires? J. appl. Psychol., 1940, 24, 758-769.
23. URITT, S. D. Checking radio's selling power. Broadcast Advertising, 1932, V, 13-17.
24. WEISSEL, H. The coincidental audience measurement. In proceedings of the Sixteenth Institute for Education by Radio. Columbus: Ohio State Univ. Press, 1946.

APPENDIX

COINCIDENTAL
TELEPHONE INTERVIEW

Date _____ Telephone number called _____

Time of call (when you begin to dial) _____

If no answer after 5 rings, check here _____

Busy signal _____; Called again at _____; Result _____

Respondent: Man _____; Woman _____; Child _____

"This is the Radio Research Bureau calling!"

1. Were you listening to the radio just now? Yes ____; No ____.

2. To what program were you listening, please? _____

"We'd like to ask you a few questions about that program, if you don't mind."

3. Compared with other _____ programs, do you think tonight's program so far is: excellent _____; good _____; or just fair _____.

(record verbatim any other information given in answer to this question)

4. What was the worst part of tonight's program? (record answer verbatim)

5. What was the best part of tonight's program? (record answer verbatim)

Conclude interview with, "Thank you very much."

Interviewer's Signature.