

RADIO AUDIENCE MEASUREMENT

BY

MATTHEW N. CHAPPELL, *Ph.D.*

AND

C. E. HOOPER, *M.B.A.*



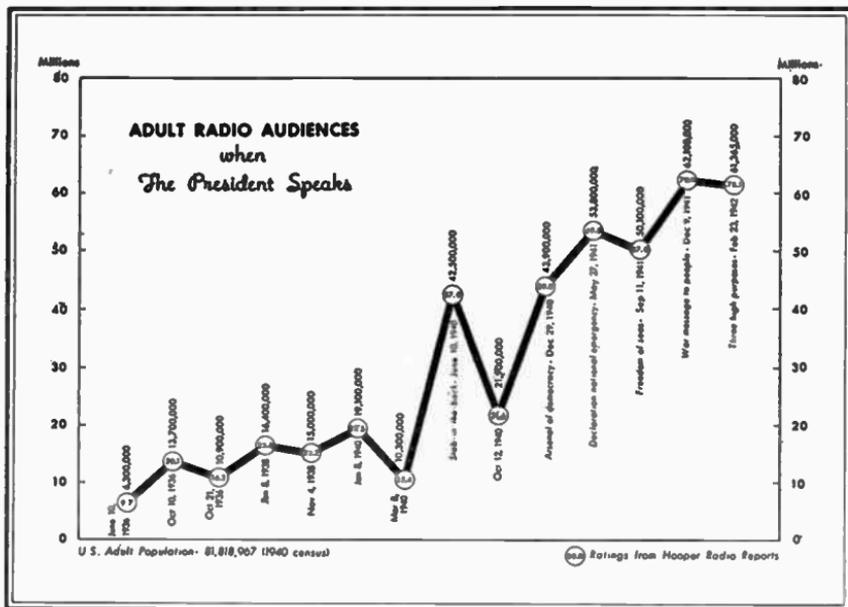
STEPHEN DAYE · NEW YORK

COPYRIGHT, 1944, BY
MATTHEW N. CHAPPELL, PH.D., AND C. E. HOOPER

ALL RIGHTS RESERVED. NO PART OF THIS BOOK MAY BE
REPRODUCED IN ANY FORM WITHOUT PERMISSION IN
WRITING FROM THE PUBLISHER, EXCEPT BY A RE-
VIEWER WHO WISHES TO QUOTE BRIEF PASSAGES OR TO
USE ONE ILLUSTRATION IN CONNECTION WITH A
REVIEW WRITTEN FOR INCLUSION IN A MAGAZINE OR
NEWSPAPER.

PRINTED IN THE UNITED STATES OF AMERICA
AMERICAN BOOK-STRATFORD PRESS, INC., NEW YORK

To RADIO'S MASTER VOICE:



The original of this chart now hangs in the White House and is destined for the Hyde Park Memorial Library

ACKNOWLEDGMENT

Some books have authors; but this one has only writers.

In the text it is remarked occasionally that the writers are responsible for one or another development. In most, if not all such cases, the original idea was forced upon us by one or more friends in the radio industry. Even the coincidental method which we have developed into radio's basic source of audience size measurement was originally presented to us by Dr. George Gallup. Many a reader whose name is unmentioned will say to himself, "I believe I gave them that idea". He did.

The glossary of radio terms which appears at the end of the book is likewise an industry job. For its existence we wish particularly to thank Philip Merryman of NBC, Francis Conrad of Edward Petry & Company, John Churchill of CBS, Sidney Fishman of MBS, Edward Evans of BNC, and Barry Ruple of NBC, each of whom is, and has been, a ready source of help and of ideas.

Among the many others on whose constant assistance we have relied and who are responsible for many of the researches we have pursued must be mentioned the following: Dr. Frank Stanton of

CBS, Samuel E. Gill of Biow Advertising Agency, Harold Webber of Foote, Cone and Belding, Charles Smith of CBS, and our good friend and tough competitor, Archibald Crossley.

P R E F A C E

It appears that we live in a revolutionary world. It also appears that we always have.

One of the many revolutions currently in progress involves the migration of research from the seclusion of the academy to the field of competitive enterprise. This movement, which encompasses most of the sciences and "social sciences" from astronomy to economics, results from several forces. One is the mounting tax rate which may eventually bring an end to privately endowed institutions of learning, science and research. A second is the growing magnitude of the problem studied—the opening of new fields of scientific endeavor.

In the study of human behavior, problems which involve an individual or even small groups of individuals may be studied readily by academicians and their graduate students. But not until business and industry developed a fundamental interest in the behavior of large numbers of people did it become possible to undertake the study of people as they live and determine the course of living.

All advertising and media research is psychological research—psychological research of such scope and magnitude that no university could

dream of conducting it. Universities neither can nor can be expected to contribute in any great measure to the science of human behavior on this scale. The responsibility for its development must rest with business.

It should be recognized that sampling work of the nature considered in this book, while it is actually the beginning of a great branch of science—the science of human behavior in the mass—is only the beginning. In this embryonic stage of development we are designing and perfecting the methods for measuring specific instances of public response to stimuli. Thousands upon thousands of these specific instances must be measured before those fundamental principles called “laws” can be abstracted from them. The isolation of “laws” lies in the future for there are as yet few, if any, isolated principles of public behavior which have such force.

While it is true that this new science of mass behavior is concerning itself almost entirely with the measurement of specific instances, the worker in this field should never lose sight of the fact that the ultimate goal is to abstract general principles which make the more wasteful measurement of specific instances unnecessary.

Charles Lamb has pointed out in rare humor that the principle of roasting meat was isolated only through repeated instances of placing a pig in a house and burning the house down. So it is with any general principle.

We are in our infancy. As we grow up, the goals—general principles—will be achieved. But first must come accurate measurements of specific instances, without which “laws” can never be abstracted.

MATTHEW N. CHAPPELL

October 14, 1943

CONTENTS

ACKNOWLEDGMENT	vii
PREFACE	ix
LIST OF CHARTS	xv
LIST OF TABLES	xvii
INTRODUCTION	xix
I. PERSPECTIVE	1
II. SELECTION OF POPULATIONS TO BE MEASURED	10
VARIATIONS IN BROADCASTS—THE STIMULI— SAMPLE REQUIREMENTS FOR NETWORK AUDIENCE MEASUREMENTS—SAMPLE REQUIREMENTS FOR RELIABLE MEASUREMENTS OF LISTENING IN A NATIONAL OR REGIONAL POPULATION.	
III. REPRESENTING A POPULATION BY A SAMPLE	25
HOW "MENTAL SETS" ARE REPRESENTED IN A SAMPLE—SOURCES OF ERROR IN STRATIFICATION OF SAMPLES.	
IV. RELATION BETWEEN SAMPLE SIZE AND STATISTICAL RELIABILITY	49
CHANCE VARIATIONS IN COIN TOSSING—RELIABLE DIFFERENCES BETWEEN RATINGS.	
V. THE COINCIDENTAL METHOD	61
DOES THE METHOD YIELD A VALID MEASURE OF LISTENING?—DOES THE METHOD YIELD A VALID MEASURE OF PROGRAM PREFERENCES?—DOES	

THE METHOD EMPLOY A UNIT OF MEASUREMENT WHICH IS CONSTANT UNDER ALL CONDITIONS?—IS THE METHOD EQUALLY APPLICABLE TO ALL POPULATIONS AND TO ALL PROGRAMS?—IS THE SAMPLE REPRESENTATIVE OF THE POPULATION FROM WHICH IT IS DRAWN?—IS THE SAMPLE SUFFICIENTLY LARGE TO YIELD STATISTICALLY SIGNIFICANT FIGURES?

VI. THE COINCIDENTAL METHOD (CONTINUED); BASIC MEASUREMENTS OF THE RADIO INDUSTRY	111
BASIC INDEXES—THE BASIC TREND LINES, EVENING—BASIC TRENDS, DAYTIME—PROGRAM TYPE TRENDS.	
VII. RECALL METHODS	138
DAY-PART RECALL—THE PRINTED ROSTER—THE UTILITY OF DAY-PART RECALL AND ROSTER METHODS—IMMEDIATE RECALL.	
VIII. FIXED SAMPLE METHODS	169
THE MECHANICAL RECORDER—LISTENER PANELS.	
IX. COMBINATION OF METHODS	206
COMBINATION OF DAY-PART RECALL AND MECHANICAL RECORDER METHODS—COMBINATION OF THE COINCIDENTAL AND PRINTED ROSTER METHODS—COMBINATION OF THE TELEPHONE COINCIDENTAL AND THE TELEPHONE DAY-PART RECALL METHOD—COMBINATIONS OF COINCIDENTAL AND FIXED SAMPLE METHODS—PERSONAL INTERVIEW COINCIDENTAL METHOD—IN CONCLUSION.	
GLOSSARY	229
INDEX	241

LIST OF CHARTS

NUMBER	NAME	<i>Page</i>
Frontispiece	The President Speaks	
I	Program History	81
II	Age and Sex Groups in Listening to "Lone Ranger"	86
III	Statistical Tolerance in Ratings	87
IV	Patterns of Listening to Charlie McCarthy	88
V	Pattern of Listening to an Audience Dissipating Program	85
VI	Pattern of Listening to Fibber McGee and Molly	86
VII	Pattern of Listening to Bob Hope	87
VIII	Audience Composition	98
IX	Use of Sponsor's Brand by Listeners and Non-Listeners to Eleven Daytime Serial Dramas	99
X	Relation of Length of Listenership to Program A to Use of Sponsor's Brand	100
XI	Relation of Frequency of Listening to Use of Brand Sponsored by Program K	101
XII	Available Audience Index—Evenings	118
XIII	Sets-in-Use Index—Evenings	120
XIV	Average Rating Index—Evenings	122
XV	Broadcast Hours Index—Evenings	124
XVI	Available Audience Index—Daytime	126
XVII	Sets-in-Use Index—Daytime	128
XVIII	Broadcast Hours Index—Daytime	180
XIX	Average Rating Index—Daytime	182
XX	Evening Sponsored Network News Programs—Average Rating	183
XXI	Evening Sponsored Network News Programs—Broadcast Hours	185
XXII	Evening Sponsored Network News Programs—Percent of Listeners	186

NUMBER	NAME	Page
XXIII	Effect of Program Age on Program Recall	142
XXIV	Effect of Program Length on Program Recall	143
XXV	Effect of Popularity Differences on Program Recall	145
XXVI	Effect on Program Types	146
XXVII	Effect on Networks	148
XXVIII	Distribution of "At Home" and "Not at Home"	150
XXIX	Distribution of Responses in the Telephone Sample in Different Seasons	152
XXX	Effect of Seasons	153
XXXI	Distribution of Responses in the Telephone Sample in Different Geographic Areas	155
XXXII	Effect of Geography	156
XXXIII	Variation of "Not at Home Then"	158
XXXIV	Effect of Time Lapse	160
XXXV	Example of a Set User's Station Tuning	176
XXXVI	Sample Recorder Tapes Show Types of Set-Users	178
XXXVII	Distribution of Evening Tuning Intervals	182
XXXVIII	Hours of Listening by Family Size	203
XXXIX	Duplication of Audience	204
XL	Memory Loss With Elapsed Time in Day-Part Recall Method	215
XLI	Flow "To" and "From" a Nationally Sponsored Network Program	221

LIST OF TABLES

NUMBER	NAME	<i>Page</i>
I	Aggregate Population Subgroups Developed by Stratification	48
II	Ten Samples of Ten Coin Tosses Each	51
III	Ten Samples of 100 Coin Tosses Each	52
IV	Ten Samples of 1000 Coin Tosses Each	54
V	Responses to the question: To what station do you listen most in the daytime?	73
VI	Responses to the question: To what station do you listen most in the evening?	73
VII	Responses based on the question: To what station did you last listen?	74
VIII	Responses to question: Who is your favorite radio personality?	75
IX	Printed Roster Recall Study in Eleven Cities	76
X	Number, Age and Sex of Listeners per 100 Listening Homes in the Evening	89
XI	Listeners per 100 Listening Homes in the Day-time Monday-Friday	91
XII	Number, Age and Sex of Listeners per 100 Homes in the Sample in the Evening	94
XIII	Sales Effectiveness of Programs in Relation to "Conscious Impression"	102
XIV	Population Subdivisions Resulting from Stratification in Eight Dimensions	186
XV	Stability of "Computed Coincidental" Ratings for Early Morning Programs	214

INTRODUCTION

The need for radio audience measurement dates from the first interview in which "time" was offered for sale. It is fundamental to commerce that the vendor be able to demonstrate how much he is offering. It is necessary to a transaction that both buyer and seller use the same measuring standard; else there cannot be the meeting of minds which is the foundation of a contract. It is inevitable, assuming its economical availability when needed, that the most precise measurement obtainable will be accepted ultimately as standard by both buyer and seller. Each of the linear, area, time, currency and other measurements which is now taught axiomatically in our grade schools has a history which involves periods of confusion when the community was making its selections from among the various concepts being advocated. Each is now treated axiomatically because the selection was completed long ago.

The buyer and seller of radio time have not yet experienced the completion of this evolutionary process. Since this is so, we consider it timely to review the methods developed thus far.

October 14, 1943

C. E. HOOPER

**RADIO AUDIENCE
MEASUREMENT**

I. PERSPECTIVE

Radio in America was relatively slow in getting under way, so slow indeed that in 1919 Radio Corporation of America was organized at the urgent request of the Navy to prevent basic patent rights from passing out of the country.¹ But despite this tardy start, American radio developed at a truly phenomenal rate. By 1940² the United States, with six per cent of the world's population, had thirty-seven per cent of all broadcasting stations and fifty-two per cent of all receiving sets. Our radio facilities and programming had become the envy of all the rest of the world.

American radio has achieved its importance primarily because it fulfills fundamental desires and needs of the American people. In this attainment, it has been guided by one fundamental principle: people are attracted to the radio only if they can get programs they want. Radio audience measurement derives from this principle. Through it, the radio industry seeks to furnish people with the programs they prefer, and not with programs which

¹ Archer, *History of Radio in 1926*, pp. 187-141.

² NBC Records—April 15, 1940. *Electrical & Radio World Trade News*, Electrical Division, Bureau of Foreign and Domestic Commerce.

some advertiser or company executive *believes* they prefer, nor yet with those which some reformer asserts they *ought* to prefer. The determination of the public's radio program preferences and desires is the basic function served by radio audience measurements.

In the early days of radio, the personal likes and dislikes of a prospective sponsor were carefully evaluated by those trying to "sell" him a program idea. One advertiser might sponsor a program because his wife liked music. Another, interested in tropical fish, might be "sold" through this hobby. One station owner actually dropped a sponsored program because his wife did not like the coffee it advertised. The board of directors of a large company wanted to drop a highly effective program because they personally did not find it interesting.

The first form of measurement used to guide programming was obtained by counting the number of letters elicited by programs. "Fan" mail count is still used for some purposes but is rapidly being replaced by more accurate determinations. It was discovered very early that people who voluntarily write letters concerning a program are atypical. They tend to be the voluble members of the fan group and are but a poor representation of all listeners. For example, it might have been judged from mail pulled by *Voice of Experience* that this program obtained a very high rating whereas its rating was quite modest in comparison

with some evening programs having relatively low mail response.

The early "measurements" used by broadcasters were no more exact than those employed by advertisers. A primary need of station management is some measure of the number of people who listen to the station habitually—that is, station "coverage". One of the first methods used for this purpose was to draw a circle on a map with a hundred-mile radius about the station and determine the number of people who live within that circle. We know now that such a procedure is entirely meaningless. Differences in power, interference between stations, overlapping of stations, local geology and geography, station programming, wave length, and numerous other factors are known to influence the size of the population habitually reached by each station.

Radio broadcasting in America has always been competitive. One of the early methods of comparing network coverage was to add up the kilowatt power of all of the stations affiliated with each network. The fact that a 50,000 watt station might be built on a geological stratum that could restrict its signal strength to a small area while a 1,000 watt station in another area built on a different geological stratum might reach out hundreds of miles was not clearly appreciated. To those currently struggling with the intricate problems of station and network coverage, these early solutions seem convenient, indeed.

One of the earliest sampling measurements in the radio audience field was made in 1929 by Archibald Crossley using a simple recall method in which he asked respondents what programs they had heard the previous day. The Association of National Advertisers and the American Association of Advertising Agencies perceived immediately the value of listening data based on sampling. As a result, they established an organization called the Cooperative Analysis of Broadcasting which designed the first continuing service for measuring programs. C.A.B. accepted the recall method used by Crossley and retained him to produce audience reports which C.A.B. sold to its subscribers on a non-profit basis.

The primary objective of C.A.B. was to furnish the buyers of radio time—the advertisers and agencies—with the means of evaluating their investments in radio.

In 1934, Clark-Hooper, Inc., measurers of magazine and newspaper advertising effectiveness, entered the field of continuous radio audience measurement using the “coincidental” method, in which respondents are asked not what they have listened to during some previous time but only what they are listening to “now”. In 1938, Clark-Hooper split into two companies, and C. E. Hooper, Inc., continued the radio audience measurements. Whereas C.A.B. was established to serve the “buyer” of radio time, the objective of C. E. Hooper, Inc., was and is to furnish audience measurements to both the

buyer and the seller of radio time. To that end, reports are published regularly not only on sponsored network, sustaining and government programs but also on individual radio station audiences in cities and throughout their total service areas.

No other audience measuring service operating on a broad scale appeared until A. C. Nielsen Company entered the field in 1943 with a method for evaluating listening through the use of a mechanical instrument which is attached to radios and which records the set operation. The purpose of this organization, like that of C. E. Hooper, Inc., is to furnish data to buyer and seller alike. The A. C. Nielsen Company regards its current operation, which covers only a part of the East and North Central areas, as experimental.

The economy resulting from radio audience measuring services in which many clients co-operate is obvious. The scope of the information produced by the operations can become very large without placing excessive financial burden on any one subscriber.

In addition to the developments made by the major audience measuring service organizations, the research of numerous individual companies has contributed to radio audience measurement. The work of the major networks has been both noteworthy and extensive. Numerous advertising agencies carrying radio accounts have established their own radio research groups to supplement the data obtained from the service organizations. Recently

one agency has established a division, the function of which is intensive research on radio commercials. This development seems destined to grow and to be extended to other research on program content.

Each method of evaluating radio audiences has its individual characteristics and necessarily produces a measurement which is different from that obtained by any other. This fundamental fact holds even in those confusing cases where the measurements are given the same name. "Program ratings" offer an excellent illustration. "Ratings" are currently offered which are obtained by the coincidental, day-part recall, mechanical recorder and printed roster methods. No program rating obtained by one method is identical with that obtained by any other. But because the term "rating" is applied to figures obtained by each of these methods, the figures are frequently thought to be equivalent. This situation has caused widespread confusion in the radio industry.

The physicist, Bridgman, has remarked that scientific concepts are defined by the operations that measure them. A "program rating" is such a concept. It is measured by some specific sampling operation and was non-existent before the particular radio audience measuring method was developed. This being the case, the "rating" cannot be understood thoroughly and used wisely in the absence of understanding of the method through which it was obtained.

How accurately are the public's radio preferences and listening behavior determined? What range of listener characteristics is measured? What are the limits of application, reliability and validity of the methods used for obtaining these measurements? Even those who have had long association with radio audience data are sometimes unable to answer these questions.

It is in hope of supplying at least a partial answer to them that the present volume is offered. The general plan of approach is to examine each of the major measuring methods in the light of each of the following six fundamental questions:

1. Does the method yield a valid measure of radio listening?
2. Does the method yield a valid measure of program preferences?
3. Does the unit of measurement remain the same under all conditions?
4. Is the method applicable to all broadcasts and all "populations"?
5. Does the sample represent adequately the "population" to be measured?
6. Is the size of the sample adequate to yield measurements which are statistically significant or reliable?

It may seem at first blush that questions 1 and 2 are identical. This is not the case. Dr. Albert Freiberg at the Psychological Corporation has pointed out that in polling or sampling operations the purpose is to obtain responses which constitute "votes" on questions of commercial or political significance.

If a respondent reports the purchase of a package of Quaker Oats or a bottle of Coca-Cola that, in a "product use" study, is a "vote for" Quaker Oats in the cooked cereal class and a "vote for" Coca-Cola in the soft drink category. And when a respondent says he is listening to Fibber McGee, that is a "vote for" the Johnson's Wax program.

As Freiberg has shown, a primary purpose of commercial sampling operations, like that of elections, is to determine public preferences by counting the "votes for" one or another form of behavior. In the field of radio audience measurement, however, some measurable behavior represents a "vote against" a program rather than a preference for it. It is therefore necessary in considering some methods to differentiate between a record of tuning and an expression of program preference.

The first four of the above questions pertain to each method as a way of collecting the information sought. The last two pertain to the sample of the population selected for purposes of study. In the pages which follow, each of the methods currently in wide use as well as some which, though new, hold great promise will be examined in the light of the above six basic questions. The objective is to establish the range of their accuracy, validity and utility.

Each method of collecting information has characteristics which apply to it alone. These characteristics will be considered in separate chapters. When,

however, the problem of sample is considered, all of the methods are subject to similar considerations. It will, therefore, facilitate matters if the problems involved in obtaining a sample of adequate size and one which represents faithfully the population from which it is drawn are considered first.

II. SELECTION OF POPULATIONS TO BE MEASURED

The term "population" is used here in its statistical sense. A "population" is any group of objects being sampled. It may or may not consist of human beings. If a forest is being studied, the population may consist of pine trees. If the behavior of radio executives is being studied, radio executives constitute the "population". If the study is made in telephone homes of thirty-two cities having "equal network opportunity", telephone homes in those cities are the "population".

Many different kinds of figures are required in the measurement of program audience characteristics and program effectiveness. Not all can be measured by the same sampling operation. Some measurements of program audience characteristics must be made in such a manner that they will be strictly comparable with similar measurements made on other programs. Among these are "program ratings". Other measurements must be made which have reference only to the client's program itself. Among these are measurements of total homes listening throughout the nation and measurements of sales effectiveness of programs.

The measurements to be made necessarily determine the population to be sampled. In order to perceive more clearly the nature of the populations that must be sampled in making different types of radio audience measurements, it will be well to consider the manner in which the stimulus—the broadcasting situation itself—may vary from program to program.

A. VARIATIONS IN BROADCASTS—THE STIMULI

There are two general types of broadcasting variables to be considered. The first may be regarded as qualitative. It consists of all those characteristics associated with a radio program which constitute its capacity to attract or recruit an audience. The second type of variable may be considered to be quantitative. It consists of all those conditions which constitute “coverage”—the area over which, by virtue of broadcasting facilities and habits of use of the facilities, the advertiser’s effort is available to listeners.

Neither the qualitative variables which constitute the audience recruiting value of a program, nor the quantitative variables which constitute coverage, are always clearly appreciated. The qualitative variables are characteristics of the advertiser’s program only. They are completely independent of the quantitative variables such as characteristics of the stations employed or the areas served or the number

of sets owned. The program is the vehicle. Coverage is where the vehicle goes.

The recruiting value of the advertiser's program is determined for the most part by the following eleven variables:

1. Time of day.

Generally speaking, radio set operation is lowest in the early morning, rises at noon, drops slightly in mid-afternoon and in late afternoon begins a climb which reaches its peak between 9:00 and 10:00 P.M. from which it declines as people retire.

2. Day of the week.

Listening is higher on Sunday evening than at any other time in the week. But the available audience (number of persons "at home and awake") is smaller than on any other evening. It appears, therefore, that the differences between evenings is at present largely a matter of programming. There are also differences from evening to evening in the composition of the available audience which have not yet been fully exploited.

3. Season of the year.

The ratings of most programs climb from September to a maximum in January or February, at which time they undergo a decline to a minimum in July or August. This seasonal decline is not always clearly appreciated, with the result that some excellent programs have been withdrawn from the air permanently when they were merely manifesting a normal seasonal decline.

The seasonal variation is much more marked for evening than for daytime programs. In the daytime, there is little change in programming from winter to summer; but the seasonal decline in evening listening is magnified by the withdrawal from the air during the summer months of many higher rating programs.

4. Content of programs.

In general, programs having high entertainment value such as variety and drama recruit the largest audiences and serious talks the lowest. However, listener's interest in program content at the time of the broadcast is the determining factor. The highest rating on record was obtained by a very serious talk—President Roosevelt's war message to the people on December 9, 1941 (see dedication chart).

5. Talent employed.

In radio, as in all other arts, the skill of the performer is a major factor in the audience's satisfaction with the performance. "Big name" talent is generally found to make a good show better. When Winchell was absent from the air on duty with the Navy, the *Jergen's Journal* rating underwent an appreciable drop.

6. The size of the audience recruited by competing stations.

7. Characteristic appeals of competing programs.

Eddie Cantor and *Fred Allen* are both variety programs which attract large audiences and which have similar appeals. For a time they were broadcast on competing networks at the same time on Wednesday evening. When *Allen* moved to Sunday evening, both programs benefitted.

Gabriel Heatter is broadcast on MBS, 9:00 to 9:15 P.M., five days a week. The competition he encounters is *Lux Radio Theater*, *Burns and Allen*, *Mayor of the Town*, *Major Bowes* and *Philip Morris Play House* on CBS; *Counterspy*, *Famous Jury Trials*, *Fitch Bandwagon* and *Gangbusters* on BNC; and *Telephone Hour*, *Mystery Theater*, *Eddie Cantor*, *Bing Crosby* and *Waltz Time* on NBC. The competition is very strong, but it is all entertainment—variety, drama, and music. None is news. As a result, 'Heatter's' news program obtains a very satisfactory rating which is about the same size each evening.

8. The size of the audience recruited by immediately preceding programs on the same and competing networks.

9. The size of the audience recruited by programs which follow immediately on the same and competing networks.

During the summer of 1948, *Fitch Bandwagon* obtained only a modest rating. *Jack Benny*, which precedes it on NBC on Sunday night, and *Charlie McCarthy*, which follows it, were both off the air. *Charlie* returned to the air in September and the *Bandwagon's* rating went up. *Jack Benny* came back in October and the *Bandwagon's* rating took another rise. Both rises were in excess of the expected seasonal gain.

Lux Radio Theater was off the air during the first week in September. When it returned in the latter half of the month, the rating of *Screen Guild Theater*, which follows it on CBS, was raised 3.5 points. At the same time, two of the three programs competing with *Screen Guild Theater* experienced a drop in rating despite the rising seasonal trends.

10. Characteristic appeals of programs immediately preceding on the same and competing networks.
11. Characteristic appeals of programs following immediately on the same and competing networks.

Bob Hope follows *Fibber McGee and Molly* Tuesday evening on NBC. Analysis shows that *Fibber* builds a big audience which he passes on to *Hope*. As a result, *Hope's* audience is already assembled at the opening of the show. This helps his rating greatly since he does not have *Fibber's* building job. But *Hope* also attracts some listeners who start tuning in during the latter part of *Fibber's* broadcast. This helps *Fibber's* rating.

This list by no means exhausts the conditions of broadcasting which affect the size of the audience recruited by the advertiser's effort. There are many others over which he has no control such as temperature, rainfall and other weather conditions, national and international events, etc. But the eleven listed

are those he selects, buys and pays for as a vehicle for carrying his message.

It is to be noted that only the variables 4 and 5 are inherent in the program itself. But the advertiser does not present "a program" in a vacuum. It is presented on a certain network, toward which certain habits of listening have been developed, at a certain time of day, during certain months of the year, in a certain framework of programming, and in competition with certain other advertisers who are doing the same thing. His success in designing an effort to recruit listeners is a result not only of how well he designs his own program and selects its talent but also of the shrewdness, insight and understanding with which he has selected his network, his competition, his position in a structure of programs, the time of day, day of the week and season of the year.

The first radio audience measurement sought by advertisers is that which will enable them to determine the success with which they have judged all of these eleven variables which determine the audience attracting value of their buys.

"Success" is, of course, meaningless except as the advertiser is capable of comparing his own with that of his competitors and every other network advertiser. The measure of the attraction or recruiting value of the network advertiser's effort is called a "network program rating". It follows that the primary and essential characteristic of "network

program ratings" is that *the "rating" on one advertiser's effort must be strictly comparable with the "rating" on another advertiser's effort.*

The demand for strict comparability from program to program is a most important element in determining the "population" to be represented in obtaining "network program ratings".

The quantitative element of the advertiser's purchase—coverage—he buys by the piece (station) depending on the areas he wishes to reach. This quantitative element may be increased or decreased without affecting the eleven qualitative variables which govern the attraction value of his effort. Technically speaking, the quantity, coverage, is an "independent variable". It, therefore, introduces the need for a second measurement of radio audiences, one which includes both the eleven qualitative variables as well as coverage. This measurement is the "cross-section rating", *whose primary characteristic is that it must reflect reliably all significant conditions influencing coverage.*

B. SAMPLE REQUIREMENTS FOR NETWORK AUDIENCE MEASUREMENTS

It is sometimes believed that both of the above measurements, "network program ratings" and "total homes" or "cross-section ratings," are one and the same; or that they may both be measured in the same sample. Consideration of the basic principles

reveals the fallacy of this belief. The measurement of "network program ratings" requires rigid exclusion of the coverage variables while the measurement of "cross-section ratings" requires their rigid inclusion. Each measurement, therefore, requires a totally different distribution of sample. Attempts to obtain both measurements from a single sample would seem to be based on a lack of understanding of the variables operating in radio broadcasting.

An illustration will make clear the difference in the samples required. Suppose, for the sake of argument, that knowledge was available which would enable a sample of a normal cross-section of radio listening behavior to be developed and that this sample of, let us say, 100,000 homes, or whatever the variables of coverage might necessitate, represents all radio homes in the nation and reflects reliably all the variations of station facilities and listening behavior that operate in the nation.

Suppose, further, that in this sample the audiences listening to programs A and B—both broadcast on the Blue Network—are measured by some meaningful and valid method; and that 12,000 of the 100,000 homes are found listening to Program A and 15,000 homes are found listening to Program B. These results show that twenty-five per cent more homes listened to Program B than to Program A.

But which program has the greater attraction value? Which is the more efficient advertising ef-

fort? Which advertiser was more successful in his selection of time, program, competition, program structure, etc.?

Data obtained from a normal cross-section of the national population would offer no answer to this question; for Program A may be carried on fifty-one stations of the Blue Network, while Program B is carried on eighty-seven, only thirty of which also carry Program A. Under these conditions, Program A could not be heard in many homes in the nation which listened to Program B and vice versa. The presence of the coverage variable, which is essential to measurement of "cross-section ratings", vitiates completely the measurements of comparative attraction values—the "network program ratings".

A figure based on an adequate cross-section sample of radio listening homes would tell one, and only one, thing—the number of homes in which the program was heard. Because it includes coverage which varies within exceedingly wide limits from program to program, it would afford the advertiser no basis for determining the comparative attraction value of his effort or for judging the shrewdness of his selection of time, network, program, talent, competition and position in programming structure.

In order to obtain "network program ratings"—measurements which are strictly comparable from one program to another, quite different conditions of sampling are required. It is necessary first to es-

establish standard conditions of measurement which will completely eliminate variations in coverage from program to program, from network to network, and from one time of day to another. *The chief characteristic of these standard conditions must be that all measurements are made in localities in which each of the four networks can be heard with equal ease in both the daytime and the evening.* The last phrase is particularly significant; for many stations which can be heard with the greatest of ease in a given area in the evening, can be heard in the daytime either not at all or to only a limited extent, and vice versa.

The selection of "areas of equal opportunity" for all networks is no simple matter. However, there are at least thirty-two large cities in the nation in which all four networks have local representation. It is among these thirty-two cities, if anywhere, that conditions of "equal network opportunity" might be expected to prevail. Even here the fact of "equal network opportunity" must be demonstrated and not assumed, because network station power in the thirty-two cities varies from 250 watts to 50,000 watts.

As a result of intensive and extensive studies of the areas of "equal network opportunity", C. E. Hooper, Inc., has selected these thirty-two cities as satisfying the standards required for developing "network program ratings" which are free from variations in coverage and strictly comparable from

one program to another regardless of program type, network used, time of day or season of the year. Under conditions where the variable of coverage is thus eliminated with a high degree of precision, and only under such conditions, is it possible to obtain "network program ratings" which are strictly comparable with each other.

From these considerations it becomes apparent that "network program ratings" must be obtained in only certain localities that are known to be areas of "equal network opportunity". The "population" which must be represented in the sample is the population of these localities only.

C. SAMPLE REQUIREMENTS FOR RELIABLE MEASUREMENTS OF "CROSS-SECTION RATINGS"

The sampling problem as it applies to "network program ratings" is relatively simple compared with that involved in obtaining a valid measure of "cross-section listening". In the former case the problem resolved itself into one of excluding rigidly all variations in coverage, and measuring only the influence of the qualitative variables.

The development of "cross-section ratings" demands inclusion of the eleven qualitative variables and also a rigid inclusion of all of the factors which influence coverage throughout the nation. The magnitude of such an undertaking is more readily perceived when it is realized that the operation must

reflect coverage in the nation at least as well as it is reflected in the network studies designed to study "coverage" alone, if precise results are required.

The term "coverage" is widely used. But it has, as yet, been but poorly defined because it is composed of so many variables. It has some relation to signal strength but the relation is far from being one of perfect correlation, as might be supposed from "coverage" maps laid out in millivolt lines.

The area over which a signal of given potential—0.5 millivolt, for example—may be thrown by a station is in itself the result of numerous variables: type, power and frequency of the transmitter; geological conditions about the transmitter, between transmitter and receiver, and about the receiver; geographical conditions between the transmitter and receiver; daylight and dark; shielding to protect other stations on same frequency; etc. Millivolt lines may be made quite meaningless either by "high noise level" or by listening habits. For example, an 0.5 millivolt signal may be clearly audible at one point but quite inaudible in another where the electrical "noise level" is high, as in cities or areas close to a power transmission system. Then, too, a given station may throw a clear signal over a given area but it may never be listened to in that area because some other station carrying the same network programs gets the audience.

The quantitative variables as they apply to a single station are complex enough. But when we con-

sider that there are currently 916 commercial stations in the nation and that approximately half of them are affiliated with one of four different trans-continental networks, that in some localities there are seventeen competing radio stations and in other localities there are none at all, the magnitude of the task of laying out a sample of homes which represents validly listening conditions throughout the nation becomes more apparent. At the present writing, BNC has 167 local outlets, CBS has 130, NBC has 141 and Mutual has 211. A month from now the number of affiliated stations of each network may be different. In no case do the cities comprising one network completely overlap with those constituting another. The signal strength of each of these network stations varies in accordance with power and frequency. (The power of network stations ranges from 100 to 50,000 watts. Frequency, whose relations to coverage is more complex than that of power, varies from 550 kilocycles to 1600 kilocycles.)

But "coverage" includes more than signal strength. Radio listening is a form of human behavior, and human beings do not distribute their behavior in rigid conformity with radio facilities. For example, the facilities of twenty-four stations may be available in a given area. One family may listen to seven of them and another to seven of which only three are heard in common, while still another listens to one station almost entirely, etc.—

throughout all the possible combinations of station listening. If a sample were to represent a cross-section of radio listening conditions, all the significant combinations of factors effecting listening behavior would have to be represented. The size of the sample which would represent the significant combinations of available facilities and the significant combinations of habits even for a single station's listening area becomes large.

Finally, if a sample which validly represented a cross-section of daytime listening conditions were developed, it might have little validity after sundown when signal strengths and habits change greatly from the daytime conditions. Another sample might be required to represent evening conditions of listening. Variations in conditions from daytime to evening may be illustrated by a single case. In the daytime, ninety-nine per cent of the listeners in a given area may be found tuned to a local 100-watter. Only a few powerful sets in the locality are able to pick up any other station at that time. But in the evening, all the networks reach the area and the particular 100-watter gets but a small share of the listening audience.

All of these variables, taken into account in a coverage study, must necessarily be taken into account in a study designed to measure "total homes listening" which is the result of the combined influence of both the program attraction variables and the coverage variables. The coverage variables have been

studied for most part by the networks. Everybody who has concerned himself with the problem will probably disagree about some of the characteristics of the methods required for measuring it, but there are two characteristics upon which all agree. The first is that the sample for measuring network coverage must be large—in the neighborhood of 100,000 homes as a minimum. The second is that every county in the nation must have representation in the sample.

It will be shown in a later section that there are at hand currently no data which enable the sampler to stratify a cross-section sample of national "populations" for competition between stations. The best that can be done now to approach proper representation of conditions of station competition is to represent the population in terms of number of local radio outlets, their type of service (network or non-network), and the combinations of different local competing network and non-network stations. While this procedure has been put in operation by the writers for the purpose of obtaining the most reliable cross-section sample it is possible to obtain at present, it must be noted that it is neither the complete nor the perfect representation of station competition.

III. REPRESENTING A POPULATION BY A SAMPLE

The fundamental assumption on which sampling operations are based is this: conditions which influence the measurements obtained are represented in a sample in the same degree to which they exist in the population studied. This is an assumption of tremendous scope, but only when it approaches a fact may a sample be said to "represent" the population from which it is drawn.

What is the nature of the factors that must be represented in a sample of radio listeners? How can anyone possibly know all the conditions which may influence the radio listening in any population? And if all of them cannot be known, how can they all be "represented" in a sample?

In the light of the basic assumption, the problem of developing a sample that represents the population from which it is drawn appears to resolve itself into a paradox. Further examination of the above three questions reveals, however, that the paradox is more apparent than real.

The conditions to be represented in a sample that is designed to measure radio listening are those

operating on each individual of the population which aid in determining how he will respond to the programs available to him.

Consider first the complexity of human responses. It used to be said that, in a baby, food in the mouth was a "native" stimulus to swallowing. By this it was meant that food in his mouth automatically led the baby to swallow it. Now a baby is the simplest human form but even he has no such invariable responses. How he will dispose of food in his mouth depends upon his past experiences and on other feelings he has at that time. If he has had no previous unfortunate experience with the food and if he is receiving from his stomach those nervous signals which are called "hunger contractions", he will probably swallow the food in his mouth. But if he receives concurrently from his stomach signals of distention which come from a fullness, the chances are good that he will spew the food on his bib. Or if he has learned through experience that he can command parental attention thereby, he may spit it out even if he is hungry. The effects of past experience together with those of other stimuli acting on him at the same time have been called by Professor R. S. Woodworth the individual's "mental set". The baby's "mental set" is quite as important as the food in determining what his response will be when the food is presented to him.

And so it is with the radio listener at 9:00 P.M. on a winter's Sunday evening when *Walter Win-*

chell is presented to him. Past experience may cause him to be "mentally set" to reject Winchell. Or he may like Winchell but prefer *Radio Reader's Digest*, *Old Fashion Revival* or *Manhattan Merry-Go-Round*, which are offered at the same time. What he listens to, if he listens to anything, depends upon those factors of training, past experience and immediate interests and activities which constitute the individual's "mental set".

The "mental set" is by no means a static condition. Indeed, its chief characteristic is flux. It changes with every situation one encounters in the course of the day, and as a result of every new experience.

Consider, for instance, how you respond to traffic noises. When you are crossing a street your "mental set" is such that you respond to them with dispatch and with specific movements. However, as you are reading this, traffic noises may also reach you. But you are now "mentally set" for reading. In this situation you react to the traffic noises very differently. You "tune them out" of your focus of observation and are nearly, if not quite, unaware of their presence.

The effects of change in "mental set" from business to home life are sometimes ludicrous. If one of the reader's female subordinates in business were to tell him that he was a moron, the reader would probably react by firing her. But if, at home, the reader's wife made the same remark, past experi-

ence would have taught that in this situation the appropriate response is, "Yes, dear".

Many of the elements composing the individual's various "mental sets" are habits, attitudes, emotions, opinions and knowledge acquired by earlier experience. For example, a factor of past experience that must be present in anyone who reads this book is that he must have acquired a primary interest in sampling or in data developed through sampling. The probability that any fifty-year-old woman who has three children and has always been a housewife could be "mentally set" to respond to this book by reading it is infinitely small.

It is apparent from the above that the elements to be represented in a sample are the "mental sets" or states of mind existing in the total population. When the states of mind present in the population are present in the sample in the same relative weight, the sample is "representative" of the population.

A. CONDITIONS KNOWN TO INFLUENCE LISTENING

It is obviously impossible for anyone to know all of the states of mind having significance to radio listening behavior which exist in any population. However, a few conditions of living and broadcasting which may influence the listening of large sections of a population are known. Most important

among these are: the availability of a program to listeners, events external to radio occurring at the time of a broadcast, competition between programs available to the listener at the same time, conditions of habit, interests and time of day resulting from geographic position of the listener, size of locality, family size and composition, intelligence, education, race, occupation, and economic status.

a. Program availability. The most important single factor in developing a sample through which to measure program listening is that all those members of the population shall be "represented" to which the given program is available. This fact is so obvious that it is easily overlooked.

No commercial program can reach all of the radio homes in America because no network has 100 per cent coverage. Furthermore, few commercial programs are broadcast over all of the stations affiliated with the network they employ. It is apparent that in the case of every program there are segments of the population who do not react to the program by listening because they cannot receive that program.

b. Time. "Mental sets" change rapidly. For example, when Fred Allen presents an original "gag", the listener may find it exceedingly funny. But having heard it, the listener's "mental set" is changed to such a degree that when he subsequently hears it on a half dozen other variety programs he may be mildly nauseated. Similarly, a program such as

that of Major Bowes (see Chart 1A) may contain a startling new element which builds a huge audience rapidly but may lose part of that audience as the newness wears off.

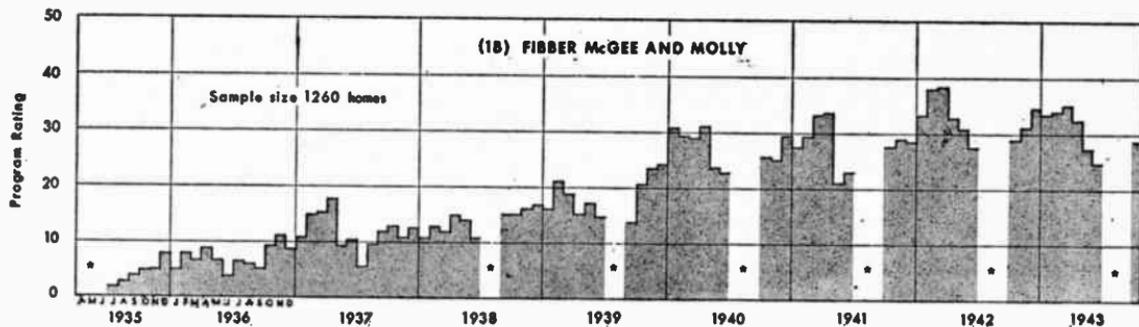
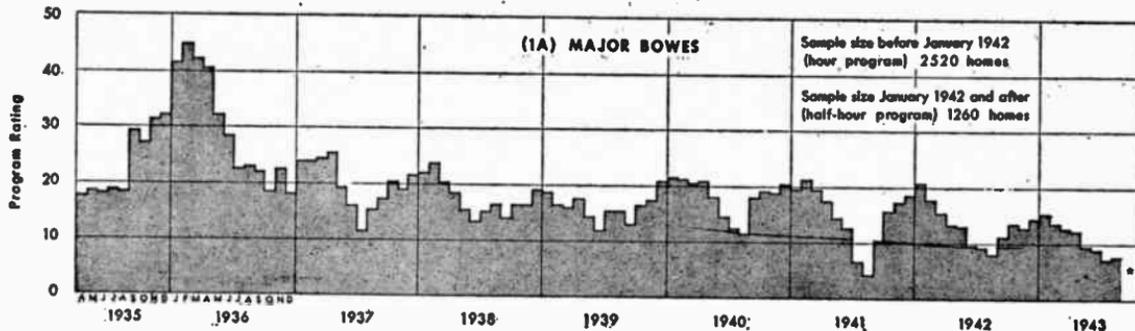
A good program which is new and with which no favorable associations have been developed in the listeners' minds obtains a smaller audience than it does after it has been on the air for some months or years and the population has had a wide experience with it. The history of *Fibber McGee and Molly*, represented in Chart 1B, illustrates this. There has been little or no change in the character of the program during the years it has been broadcast. There is nothing spectacular or startling about the show. It is just an exaggeration of all of us. Year by year it has increased its audience until, at the current writing, it consistently rates at or near the top.

The importance of the time at which audience characteristics are measured can be perceived the more readily by examination of the trend charts in Chapter VI. These show wide variations in the characteristics of the radio listening audience from one season of the year to another and from year to year. The influence of seasonal variations on the activities of a population and the effect of world events on radio listening will be considered in detail in that section. Individual sampling measurements have significance only for the specific time at which they are made. But in radio programming one requires

CHART I
PROGRAM HISTORY.

Each rating is the percent of the total homes called which reported listening to the subject program.

Source: Hooper Monthly Ratings



*Off the air

data that have wider significance. It is necessary, therefore, to make the same measurements at many different points in time, thereby developing trends of listening characteristics which give the measurements meaning over a wide range of times.

Trends, not isolated "one shot" measurements are required in radio audience study. For example, an unsuspecting audience measurer may be asked by a station owner to come into his city and make a careful, well-controlled "one shot" study of listening to each radio station in the city. The study may be made and it may show that the client's station is way out ahead of the others.

The unsuspecting audience measurer may be bearing false witness if he assumes that his findings have any meaning other than for the time at which he made them. Possibly, quite unknown to the measurer, the station owner has "plugged" his station strenuously with newspaper and other promotion for a few weeks just prior to the time when the unsuspecting one made the measurement. The costly promotion may have resulted in high listenership to the station during the time in which the measurement was made and may have been dropped immediately after. Situations of this kind are not entirely unknown in radio station audience measurement; but continuous measurements which produce listening trends covering long periods of time make such prestidigitation too expensive.

c. Competition between stations. In areas in

which a program is available to the population, one of the chief factors in determining the size of audience it will recruit is the degree of competition with other radio stations which it encounters in each locality or home. In a village of 6000 population near New York City, the signals of twenty-four radio stations are clearly audible for listening without interference from any other station in the daytime. This figure is exclusive of short wave. In another village of about the same size located at a distance from any metropolitan area, a signal of adequate strength may be heard from only one station.

An example will illustrate the importance of competition between stations. A daytime program broadcast over a wide area received a rating of 1.0 in cities which had local outlets for all four networks. In areas in which only the station carrying this program (and in some instances one other station) could be heard, the program obtained a rating of 16.0.

Similarly, the program *This Is War* obtained a rating of 20.7 in cities where it was carried by the stations of all four networks; but in a city in which it was carried by only three networks, and the fourth network station carried a commercial program, *This Is War* obtained a rating of only 8.3. The range of listeners' choice—the extent of competition between programs—is one of the major factors in determining audience size. It is also a factor which is sometimes overlooked entirely by

investigators purporting to measure radio audience characteristics on a national or regional scale.

d. Geographic conditions. Experience indicates that geographic distribution is also an important population variable influencing listening. Local time varies from one geographic area to another and causes major differences in listening. For example, if it has no "rebroadcast", a network dance program broadcast in the East at 9:00 P.M. reaches the Pacific Coast at 6:00 P.M. when even the jive artists have other interests.

Weather also varies widely from one geographic area to another and influences audience sizes. Radio measurements based on listening in homes are influenced directly by the size of the "not at home" segment of the sample. In the winter there may be fifty per cent more homes in which no one is at home in the Pacific area where weather is milder than is the case in the East where winters are more rigorous.

Tastes in programs also vary from one geographic area to another. Gene Autrey's appeal to the South-West is greater than to the East. *Pot O' Gold* appealed to the South much more strongly than to any other geographic area.

e. Size of locality. Size of locality may also play an important part in determining responses of a population. To understand clearly the importance of the size of the locality in which a listener lives it must be noted that, in recruiting listeners, radio is

in competition with all other activities in the individual's environment. A wide range of stimuli compete with radio for the individual's attention—movies, theaters, lectures, clubs, societies, newspapers, magazines, store bargains, exhibits, etc. As the city size decreases, the volume of such competing stimuli also decreases.

To take but a single example, a "big name" news reporter may get a rating of 10.0 in cities where there are available several newspapers which have good coverage of both foreign and domestic news. In small cities which have no newspapers with good foreign and domestic coverage, this radio news reporter may get a rating of 20.0.

Then, too, the competition between radio stations is maximum in large cities which have several local outlets and tends to decrease with city size. As a result of the decreased station competition, those programs carried in the smaller localities tend to recruit larger audiences.

f. Family composition. The importance of family composition in determining the "mental sets" of radio listeners is immediately apparent. Listening in a family composed of one man or one woman or a man and wife would be expected to vary widely, both in amount of listening and in selection of programs, from a family composed of a grandmother, mother, father and four children.

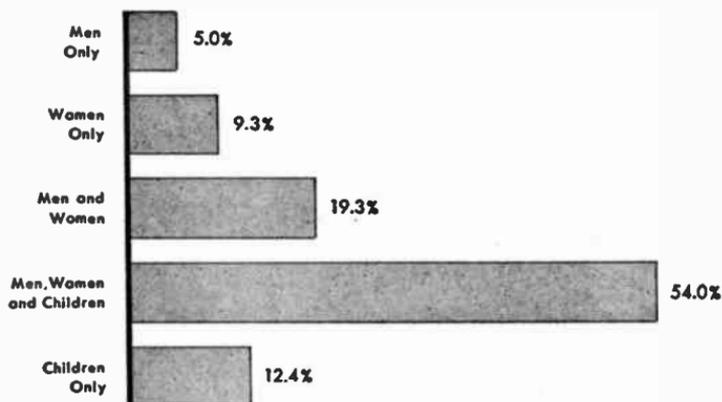
Chart II shows that *Lone Ranger* is heard in some homes in which men only, women only or men

and women only, are listening; but in about two-thirds of the listening homes children are listening and in more than one-half of the listening homes, adults are listening with them. Many a parent has

CHART II
**AGE AND SEX GROUPS IN
HOMES LISTENING TO "LONE RANGER"**

MAY - AUGUST, 1943

100% = 363 homes listening to "Lone Ranger"



been converted into a *Lone Ranger* listener because a young son or daughter tunes in the Blue with clock-like regularity on Monday, Wednesday and Friday at 7:30 P.M., E.W.T. The presence on the air of *Blondie*, *Easy Aces* or the *Roth Orchestra* may be completely unknown to the parents of the *Rangers*.

It is also clear that the amount of time when

someone in the family is listening to the radio is related directly to the number of members in the family. (See Chart XXXVIII, page 203.)

g. Intelligence. The intelligence of the population varies within wide limits and, of necessity, is an important factor in determining the individual's interests and activities. Unfortunately, there is available no practicable way of measuring the intelligence of the members of a sample or of a population which the sample represents. The best that can be done in most sampling operations is to use as a crude index of intelligence some crude measure of education such as the last school grade attended.

Based on schooling, differences in listening are less sharp than is generally supposed. For example, it is frequently said that daytime serial dramas are listened to only by people with no education. This is entirely erroneous. Leda P. Summers' study of daytime listening made in 1942-43 shows that serial listeners are well distributed throughout the economic and educational levels. Similarly, in the study of daytime serial listening in wartime made by C. E. Hooper, Inc., in Detroit, 25 per cent of the listeners in upper socio-economic level, 35.3 per cent of the listeners in the middle level, and 32.2 per cent of the listeners in the lower level were found listening to serials. These findings are further confirmed by the Blue Network's 1943 study of daytime listening.

h. National origin. Because America is the "melting pot" in which there is a large segment of first and second generations from foreign lands, national origin and race becomes significant to radio listening habits. The culture in which one has been reared together with the emotional strains and stresses influencing minority racial groups produce "mental sets" which are of major significance in determining responses to a wide range of stimuli, including radio programs. Race is particularly important to listening habits in those areas in which foreign language and culture groups are concentrated and in which stations broadcast programs in foreign tongues. In Hamtramck, the Polish section of Detroit, the programs and stations preferred are quite different from those chosen a few miles away in the Grosse Point section where national origins are also different.

i. Occupation. People are interested in those matters most immediate to their lives. Occupation is a factor of primary importance in determining the habits and interests of most Americans. From the standpoint of radio broadcasting the most important occupation by a wide margin is that of the housewife. Examination of the sponsored network program structure for daytime will reveal that practically all network advertisers are interested in housewives primarily.

The daytime serial drama, which is perhaps one of the most important literary forms that has ever

been developed, exists solely because housewives find it entertaining.¹

j. Economic status. Closely associated with differences in intelligence, educational achievement, race and occupation are differences in economic status. In fact, all too frequently in polling research casual relation is attributed to economic status when in actual fact the true cause is one of the above associated factors. It is extremely difficult to obtain any objective index of economic status which remains meaningful from one area to another, from one occupational group to another or from one time to another. In practice, the economic status of the family is usually determined intuitively by the interviewer who passes judgment on *the standard of living* in the home based on its examination and the interviewer's knowledge of the local conditions of living, and on the social, educational and cultural level manifested.

Objective factors used commonly in judging economic status are: rent, size and type of dwelling, car ownership, telephone ownership, type of refrigeration, and currency income. But in war years, none of these have much significance to economic status. Maximum employment, high wages for workmen, lack of consumer goods and food ration-

¹Daytime serial listeners have led Dr. Paul Lazarsfeld to believe that they listen for educational and psychotherapeutic purposes; but the writers have found that if people are asked why they listen to programs that give away money or other prizes they also say, "It's educational."

ing make the relation between income and habits and interests almost meaningless. A family whose home indicates a relatively low standard of living may have four males and one female gainfully employed and have a family income of ten thousand dollars or more. Such families are currently by no means rare in cities devoting themselves with enthusiasm to the industrial side of the war effort.

B. HOW "MENTAL SETS" ARE REPRESENTED IN A SAMPLE

The above list of factors which contribute to determining the "mental sets" of the potential radio audience is by no means exhaustive. It includes only ten known factors among many thousands, mostly unknown, which determine whether or not one will listen to the radio at a given time and to what program he will listen.

This brings us back to the apparent paradox of sampling. "How is it possible to represent in a sample, with their proper weight, the thousands of factors which may be important in determining listening, when most of those thousands of factors are quite unknown?" The answer to this question lies in the random selection of the sample.

A pure random sample is one in whose development every member of the population represented had just as much chance of being selected as every other member of the population.

When all members of a population have exactly the same chance of becoming members of the sample, every "mental set" which may influence the listening behavior has opportunity to be represented in the sample with its proper weight regardless of the fact that its significance to listening may be quite unknown to the sampler.

The difficulty with using a pure random sample of a population is that it is exceedingly burdensome to obtain in most sampling operations. Consider the requirements for obtaining a random sample which would represent homes in the nation. First it would be necessary to get the addresses of all the homes. This might be done in conjunction with the U. S. Census Bureau. These addresses would then be written on cards and all of them thrown in a giant mixer and mixed until they were in a thoroughly random order. Then, at random intervals while the mixing continued, enough of these cards would be drawn off to yield the sample size required. Or each address might be given a number and the sample selected by having recourse to tables of random numbers. Through either of these procedures, each of the families in the country would have an equal chance of being selected. But obtaining all the addresses would be somewhat onerous and costly. Though, theoretically, it is the most satisfactory sample that could be developed, the pure random sample is rarely, if ever, used in commercial research.

Consider what would be found in a pure random sample of homes of the nation. For every Pacific Coast home that turned up in the sample there would be about three North Central homes and five Eastern homes. But it was already known from the 1940 U. S. Census that this was the relative geographic distribution of homes in the nation. The knowledge obtained from the Census makes it unnecessary to start random selection with a national sample. Instead, the sample may be "stratified" for the geographic distribution of homes. That is to say, the number of homes to be selected in each geographic area would be decided beforehand and the random selection of homes might then be started with the geographic areas themselves. But this, too, is unnecessary, because analysis of the Census figures also reveals the proportions of the population of each area which live in large cities, small cities, towns, villages and rural communities. Therefore, instead of starting to develop a random sample of a whole geographic area, each area might be "stratified" for size of locality. From the Census figures, it could be determined how many homes must be chosen in each of the sizes of locality in each of the geographic areas. Then the sample in each locality of each size could be selected at random.

By "stratifying" a sample, random sampling is applied not to the total sample but to small subdivisions of it. At the same time, "stratifying" multiplies the number of population subdivisions in

which pure random samples must be obtained. "Stratifying" the sample is the process of applying to the sample construction known population characteristics before the random selection of homes in the multiplicity of population subdivisions is undertaken.

The calculation of the number of population subdivisions in a stratified sample is a simple mathematical process. Suppose, for example, that it was desired to stratify a sample for the following: geographic area, number of stations habitually heard, size of locality, economic status and family size. The aggregate number of population subdivisions increases with each condition for which the sample is stratified as may be seen from Table I.

TABLE I

Aggregate Population Subgroups Developed by Stratification

Condition Stratified	Number of Subdivisions	Aggregate Population Subgroups
Geographic Area	5	5
Number of Stations Habitually Heard	6	30
Size of Locality	5	150
Economic Status	4	600
Family Size	5	3000

The table shows that as the number of conditions for which the sample is stratified increases, the number of population subgroups increases as the multi-

ple of the subdivisions of the condition stratified. If the sample were stratified for only the first three conditions, there would be only 150 population subdivisions within each of which random samples would have to be obtained. By adding the last two, the number of such subdivisions is multiplied by twenty.

The number of conditions for which the sample is stratified bears an important relation to sample size, for the smallest number of homes that can represent the smallest population subdivision is one family. Further, the size of each subdivision is different. The largest would be some hundreds of times larger than the smallest where five conditions are stratified. For example, in each geographic area there would be about six times as many homes of C economic status as there would be of A status. And each of the subdivisions for the East would be twenty times as large as the corresponding categories for the Mountain area just on the basis of geographic distribution of homes. And so on, for each category of each breakdown.

The determination of the relative size of each subdivision may seem to the lay reader to be an almost impossible task, but is comparatively simple if the relative weight of each category of each breakdown is known. However, because the sample size expands so sharply with each added condition stratified, samples rigidly stratified for more than three population variables are purely scholastic. They

are impossible within the practical limits of radio audience sampling costs.

It will be shown below that the percentage of families in the nation which habitually listen to one, two, three, four, five or more radio stations is quite unknown. Therefore, in actual practice, no stratification for competition between stations can be developed at the present time. The conditions required for obtaining data which will make this most important stratification possible will be discussed in a later section.

C. SOURCES OF ERROR IN STRATIFICATION OF SAMPLES

The assumption underlying the stratification of a sample is that it will yield results which are slightly more reliable than would be the case with a pure random sample. Under the best of conditions, this assumption is warranted, but in many sampling operations attempts at stratification result in distortion of the results, sometimes to a serious degree. The two most important sources of error arising in stratified samples result from the following: (1) use of obsolete data on population characteristics; and (2) failure to apply randoming in selecting the members of large subdivisions, such as the towns composing a given division of size of locality.

a. Influence of obsolete data. It is possible to stratify a sample for a specific variable condition

only when there are at hand data from previous study which reveal the distribution of that variable in the population. Stratification for the geographic distribution of homes is possible because the U. S. Census has shown how homes are distributed throughout the nation. So with income, family size, and all of the other conditions for which stratifications of commercial samples are made; the distribution of each must have been determined by previous study and published in reference books for use in business and sampling.

In periods in which a population is relatively static, published data may serve very well for sample stratification. However, in a period during which the population changes rapidly, such as that between 1940 and 1944, figures on geographic distribution and other characteristics of a population may be obsolete before they are published. For example, any study made in 1943 based on previously published income figures would have but little meaning because of the rapid growth of employment and employment shifts from service to manufacturing industries, with resulting high wages, etc.

Similarly, in the year 1943 stratification of a sample in terms of telephone and non-telephone homes might lead to serious errors. The published figures available for stratifying a sample for telephone and non-telephone homes in Detroit in 1943, for example, would have shown that forty-seven out of each 100 homes had telephones. However, this

figure was developed in 1940 and so rapid was the change in distribution of telephone homes during this period that, by June 1943, sixty-three out of each 100 homes in Detroit had telephones. It is clear from this that any sample stratified in accordance with the 1940 figure might result in gross and costly errors in 1943.

When a situation arises in which the sampler has good reason to believe that the most recent stratification data are obsolete, it becomes necessary to drop that particular stratification of the sample and resort to random selection to yield a proper distribution of the particular population characteristic.

b. Random selection of members of subdivisions. An error that is frequently made by samplers not too familiar with stratification of a sample is that of failing to apply randoming even in the selection of members of the sub-groups of the conditions stratified. For example, 878 towns in the United States have a population between 10,000 and 50,000. In few, if any, sampling operations would all of these towns be sampled. Some, like Scarsdale, New York, are suburbs of large cities and have available facilities at many stations. Others, like Yakima, Washington, lie far away from any larger city and have available to them the offerings of only a few stations.

Since the size of a program's audience depends, among other things, upon the amount of radio com-

petition it has in each locality, the selection of towns to represent the cities of this size becomes a matter of primary importance. Only if each town has equal chance of being selected for sampling can a representative selection be assured. In practice, however, it is frequently found that when the towns sampled in a study are plotted on a map, they cluster about larger metropolitan areas from which they can be studied with greater convenience to the sampler. Towns of the same size lying at a distance from large cities, which are themselves metropolises, may be excluded completely in the selection of towns.

Because of the relation between station competition and proximity to large cities, such a selection of towns distorts radio listening results.

In those towns in which five or more stations are listened to habitually the station competition is keenest. These localities are prone to lie close to large metropolitan areas, while those where one, two, three, or four stations are listened to habitually are more prone to lie at a distance from the large metropolitan areas. Since the size of the program's audience depends, among other things, upon the amount of radio competition it has, selection of towns close to a large metropolitan area gives a completely false picture of what is happening in these towns as a group. Such a condition is avoided by random selection of the towns in which interviewing is to be conducted.

IV. RELATION BETWEEN SAMPLE SIZE AND STATISTICAL RELIABILITY

Business men are prone to shy at the mention of "statistics". And little wonder. It has been remarked by Lin Yutang that for any Chinese scholar to talk in language other than that used by the man in the street would be unthinkable. In our own culture the situation is quite different. As soon as one of us gains a little knowledge in some special field he is no longer able to talk to the rest. A greater monument to the stupidity of our methods of education would be difficult to conceive.

Few scholars or technically trained people really know anything that cannot be explained in words of two syllables when the one who knows is not too infernally lazy to learn the rudiments of conversation. Technically trained people often appear to defend their prestige by using, in the colorful language of Max Wylie, "words what us poor ignorant folks don't hardly know the meaning of them". This is particularly true of statisticians.

Not only is the language used by statisticians lacking in meaning to one not statistically trained,

it is oftentimes positively misleading. For example, when a sampler reports that a rating of 5.0 has a "probable error" of 0.6, what does he mean? The layman is prone to believe that this statement means that the measurement is accurate within a margin of error of 0.6 more or less than five. But that is not what it means at all. What is actually meant is that if 100 measurements of the program were made on samples of the same size, fifty of the measurements would fall somewhere between 4.4 and 5.6. The other fifty measurements would fall beyond these limits, and some of them quite a way beyond.

The terms "standard error" or "standard deviation" (they are two names for the same thing) have a similar meaning. If a rating of 5.0 had a standard error of 0.6, it would mean that in 100 measurements made on the program with samples of the same size, sixty-eight of the measurements would fall between 4.4 and 5.6. The other thirty-two would fall outside this range.

The important thing to be known about any sampling figure is this: How *much* larger or how *much* smaller must another figure obtained by sampling be before it is *reliably* larger or *reliably* smaller than the figure obtained. This is what the statistician means when he talks of a "statistically significant difference" between results obtained by sampling. It is the only measure of statistical reliability which can have the slightest significance to an interested layman.

A. CHANCE VARIATIONS IN COIN TOSSING

There is nothing difficult involved in gaining a clear understanding of statistically reliable differences between figures obtained in sampling operations. In fact, most Americans have already learned the basic principles involved through tossing coins or, on a more complex level, by shooting craps.

When a well-balanced coin is tossed, what is the chance that it will turn up a head? It is 50-50. The chance that it will turn up tail is just as great as that it will turn up head.

But if a coin is tossed ten times, it does not always turn up five heads and five tails. In some samples of ten trials, heads may turn up only twice and tails eight times; in others the reverse may be found. The accompanying tables have been constructed from actual coin tossing. They will serve well to

TABLE II

Ten Samples of Ten Coin Tosses Each

Sample No.	Heads	Tails
1	6	4
2	2	8
3	6	4
4	6	4
5	6	4
6	5	5
7	5	5
8	5	5
9	7	3
10	3	7

TABLE III

Ten Samples of 100 Coin Tosses Each

Sample No.	Heads	Tails
1	51	49
2	49	51
3	51	49
4	51	49
5	42	58
6	55	45
7	60	40
8	39	61
9	51	49
10	48	52

illustrate most of the important factors involved in understanding the meaning of a statistically significant difference between figures obtained in any sampling operation.

That the chances of tossing a head or a tail with a balanced coin are fifty-fifty is already known. But suppose this was not already known and that it had to be determined. It would become necessary to resort to a sampling study.

a. Sample size—10 tosses. If we were to make only one measurement using a sample of ten tosses and were to believe implicitly that the resulting figures were the "true" measure of the probability of tossing a head, we would conclude (Sample 1, Table II) that the chances were six in ten; and that the chances of tossing a tail were four in ten.

But if we took ten samples of ten trials each, we would find that the chances of tossing a head are

six, two, six, six, six, five, five, five, seven and three in ten respectively. Obviously, there is something wrong. Each of these figures cannot be the "true" measure of the probability that heads will turn up. There can be only one figure that represents the "true" probability. Since with different samples of the same size we have found that the figures which measure the probability that heads will turn up is found to vary from two to seven chances in ten, and since the figures obtained from each sample measure the same thing (our chances of tossing a head), it is necessary to conclude that there is no reliable difference between figures which range from two to seven when samples of ten trials are used in coin tossing.

b. Sample size—100 tosses. The significance of figures obtained by sampling depends upon the size of the sample used. Suppose, for example, as in Table III, ten samples of 100 tosses each are used to measure the probability that heads will turn up. The first sample would show the chances to be fifty-one in 100. Subsequent samples, however, show that they are forty-nine, fifty-one, fifty-one, forty-two, fifty-five, sixty, thirty-nine, fifty-one, and forty-eight chances in 100 respectively. But again, since all of these figures are measures of the same thing, it must be concluded from these data that there are no real differences between them. That is to say, with samples of 100 trials of coin tossing, a figure of sixty is equivalent to a figure of thirty-

nine as a measure of the "true" chances that heads will turn up. The spread, thirty-nine to sixty chances in 100 is pretty wide, but it is not so wide as that obtained from samples of ten trials each, which was from two to seven chances in ten or from twenty to seventy chances in 100.

c. Sample size—1000 tosses. When we used samples of 1000 trials each (Table IV) the limits of chance variation become still narrower. The ten

TABLE IV

Ten Samples of 1000 Coin Tosses Each

Sample No.	Heads	Tails
1	497	503
2	491	509
3	521	479
4	482	518
5	500	500
6	467	533
7	466	534
8	529	471
9	509	491
10	511	489

samples show that the chances of tossing heads are 497, 491, 521, 482, 500, 467, 466, 529, 509, and 511 chances in 1000 respectively.

Again, since they all measure the same thing, it must be concluded that with samples of 1000 each, a figure of 529 is not significantly larger than, but actually equivalent to, a figure of 466. However, the range of chance variation is now much smaller

than was the case with the samples of 10 and 100.

If the results obtained from each of the three sample sizes are all converted into "chances in 1000", the decrease in the range of chance variations with increase in sample size is more readily apparent. The data show that with ten samples of ten trials each, measurements of the probability that heads will turn up may be expected to range from two to seven chances in ten, or 200 to 700 chances in 1000, a range of *500 chances in 1000*.

The ten samples of 100 trials each show that the measurements may be expected to range from thirty-nine to sixty chances in 100, or 390 to 600 chances in 1000, a range of *210 chances in 1000*. The samples of 1000 trials each show that the measurements may be expected to range from 466 to 529 chances in 1000, a range of only *sixty-three chances in 1000*.

No matter how large a sample we might take, there would be some range within which the chances of turning up heads would vary. The larger the sample, the smaller becomes the range of chance variation to be expected—the range over which one sampling figure is not "significantly different" from another.

B. RELIABLE DIFFERENCES BETWEEN RATINGS

Exactly the same conditions of chance variation apply to all sampling measurements. If a program

rating were to be measured simultaneously in ten samples of the same size, ten different ratings might be obtained.

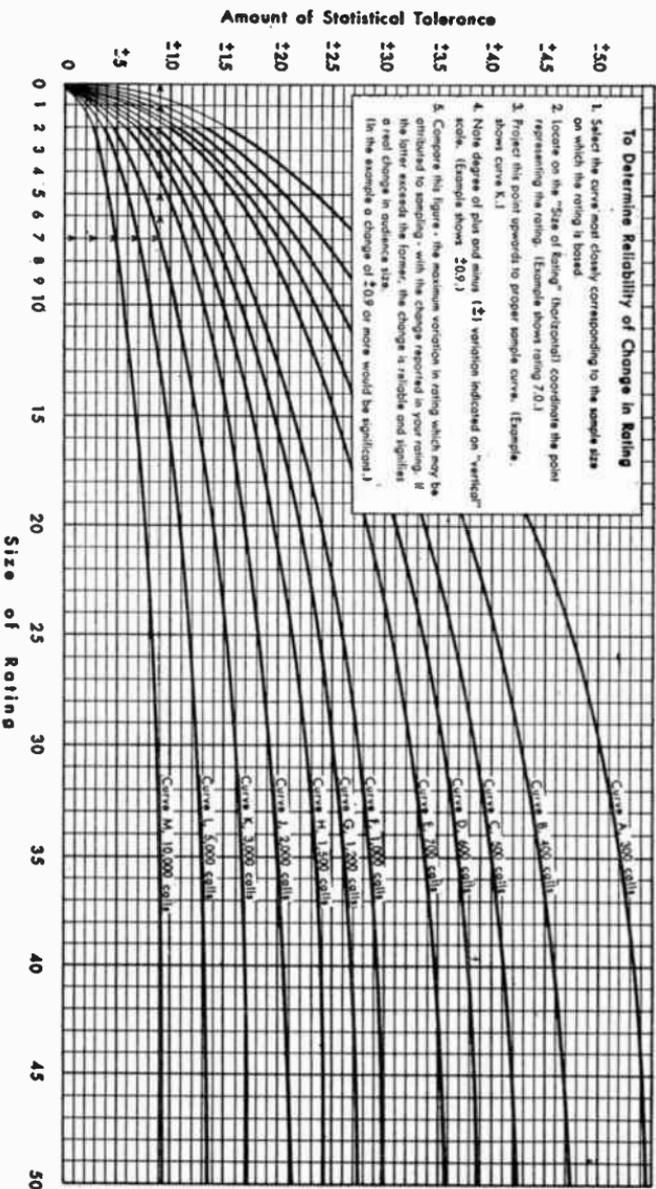
Since they would all measure the same thing, their spread would indicate the range over which one rating figure for the program was not significantly different from another.

The chart. Chart III is constructed for convenient determination of the range over which changes or differences in published program ratings are not statistically significant. While the chart is constructed for program ratings, it is equally applicable to all sampling figures expressed in percentage terms. From it can be determined the amount by which one sampling figure must differ from another in order to indicate a definite difference in size.

In this chart, program ratings are represented on the horizontal scale and the statistical tolerance ("significant differences") are represented on the vertical scale. The chart is made up of curves drawn from sample sizes ranging from 300 to 10,000. These curves may be used to determine significant differences for any samples within this range. It is impracticable to draw curves for every possible sample size. However, for all practical purposes that curve may be used which corresponds most nearly to the size of the sample on which the measurement is based.

To illustrate the use of the chart, assume that a given one-quarter hour, five-a-week program based

CHART III STATISTICAL TOLERANCE IN RATINGS



These are the variations which are unacceptable in any figures obtained by a sampling operation. The variations are expected to be less than the tolerance in 90 of every 100 samples.

These curves are drawn at 1.35 sigma of the difference. They show the maximum statistical variation expected in 91 of every 100 samples.

on 3375 interviews obtained a rating of 7.0 one month and a rating of 9.0 the next month. Has there been a real change in the size of the audience or may the difference be caused by the chance variation inherent in any sampling operation?

Since the sample on which the rating is based is 3375 homes, for all practical purposes the curve drawn for samples of 3000 (Curve K) may be used. Having determined the curve to be used, the statistical tolerance or reliable difference may be found by the following steps:

1. Locate the point representing a rating of 7.0 on the rating (horizontal) scale.
2. Project this point up to Curve K.
8. Project the point on Curve K over to the vertical scale and read off the "statistical tolerance" (significant difference).

It will be seen from the illustration on the chart that this yields a figure of approximately 0.9 on the statistical tolerance (vertical) scale. This means that any rating which is 0.9 higher (or lower) than a rating of 7.0 is reliably higher (or lower). Since the difference between the two ratings considered in our illustration is 2.0, which is greater than 0.9, the difference may be accepted as reflecting real change in audience size.

If, on the other hand, the two ratings had been 7.0 and 7.5, the difference between them would not

have been "statistically significant".¹ They would be regarded as equal to each other, indicating no change in audience size from one month to the next.

It will be seen from the chart that as the sample size decreases, the size of the maximum expected variation increases at an increasingly rapid rate.

One further fact should be noted concerning statistical variation. It pertains only to the size of the sample and has no relation to the magnitude of the error that may arise from faulty method. For example, a method may be in error by as much as twenty per cent in determining the size of a program audience in a given population. This error could be determined by measuring the audience size simultaneously by some other method of greater accuracy. However, this error of twenty per cent resulting from the method used would have no relation to the size of the *statistical* error. The latter might be only five per cent. The five percent is the error which will be present when the method is perfect—and no method is.

The statistical error is that which results from the operation of chance. It is related to sample size

¹ Statisticians should note particularly that the quantity plotted in these curves is neither the probable error nor the standard error, which can have no possible utility for the layman. The quantity plotted for each sample size is:

$$T = \frac{\text{Difference}}{\text{Sigma of Difference}} = 1.85$$

$T = 1.85$ was selected because it is the level at which the chances are ten to one that all variations are included.

only and is unrelated to the method of study employed. For this reason, it is necessary to examine carefully the method used to collect data in any study. Only by careful examination, "controlled" experiments, and direct comparison of results obtained from the same population by the different methods, is it possible to understand the sources of error that may reside in each method.

In the pages which follow, this will be undertaken in the field of radio audience measurement for the coincidental, day-part recall, printed roster and mechanical recorder methods.

V. THE COINCIDENTAL METHOD

The coincidental method gets its name from the fact that all interviews used in determining the characteristics of the listening audience are obtained while the program is being broadcast. It was not the first method for obtaining program audience measurements to make its appearance. Both the unaided recall which was changed to day-part recall and the aided recall, or printed roster, preceded it. It is, nonetheless, the most accurate and most generally desired of the methods for obtaining audience measurements, both for individual programs and for stations.

The coincidental method, developed for the most part by C. E. Hooper, Inc., has recently been adopted by the Cooperative Analysis of Broadcasting, but because the C.A.B. experience is so recent, the method will be discussed here as it has been used by C. E. Hooper, Inc. The method has many outstanding advantages which appeal to research men. First, being an interview, it gets data from direct contact with the respondent. Secondly, it is a short simple interview which requires a minimum of effort on the part of the respondent. Third, it has

a fundamental characteristic found in no other method—it measures the available audience (the “at home” element) during the time each program is on the air. Its advantages and disadvantages will be discussed at length in the pages which follow.

The coincidental questions used by C. E. Hooper, Inc., are the following:

1. “Were you listening to the 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?”

The method is currently used to furnish the following data for advertisers, networks, stations and agencies.

- | | | |
|--|---|---|
| <ol style="list-style-type: none"> 1. Size of “Available Audience”. 2. Hours of listening in “total homes” of sample. 3. Hours of listening in homes of the sample available for listening. 4. “Sets-in-Use”—per cent of total homes in the sample found listening to radio. 5. “Recruiting Efficiency”—per cent of the available audience actually listening to radio. | } | Analysed by: <ol style="list-style-type: none"> 1. Year 2. Month 3. Day 4. Hour 5. Geographic Area |
| <ol style="list-style-type: none"> 6. “Network Program Ratings”—the measure of a network program’s attraction value. 7. “Network Urban Cross-Section Ratings”—the program audience size in cities of 25,000 or more population where it is broadcast locally. | | |

8. "U. S. Urban Cross-Section Ratings"—per cent of homes listening to program in all cities of 25,000 or more population.
9. "Per Cent of Listeners"—per cent of the listening ("Sets-in-Use") audience which is listening to a specific program.
10. The internal pattern of listening to the program by one minute, three minute, or five minute intervals.
11. The composition of the audience in terms of number, age and sex of listeners per set.
12. City-by-city program ratings in terms of total homes in the sample.
13. City-by-city influence of competing, preceding and following programs on specific program rating.
14. Long-term program audience trends by type of programs.
15. Comparison of client's program with programs of same type.
16. Sales effectiveness of program.
17. Frequency of listening.
18. Length of listening in years and months.
19. Flow of audience to the program.
20. Flow of audience from the program.

} Obtained from
telephone methods
supplementing the
coincidental.

The facts in items 1 to 15 are contributed by the coincidental method alone. Those in items 16 to 20 are developed through telephone techniques described below and designed to be used in conjunction with its coincidental method. In Chapter IX, it will be shown how the coincidental method is used in conjunction with other methods to obtain listening data by size of locality, economic status and other population subgroups.

The accuracy, validity and reliability of the coin-

cidental method may be tested by applying to it the six basic questions considered in Chapter I.

A. DOES THE METHOD YIELD A VALID MEASURE OF LISTENING?

The first requirement of any method for measuring radio audience size in any selected population is that it will reflect accurately 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 more apparent when other methods are considered. In the case of the mechanical recorder (Chapter VIII), for example, behavior is measured directly. But the behavior measured is set-tuning not listening. Listening behavior is inferred but not measured in the recorder methods. No inference is involved in the coincidental method. There can be no appeal from the respondent's report. It is, therefore, difficult to conceive of a method that could measure listening behavior with greater validity.

B. DOES THE METHOD YIELD A VALID MEASURE OF PROGRAM PREFERENCE?

The purpose of measuring program audience size goes beyond the mere question of behavior. The

object is to determine listeners' preferences in programs. There is a very real difference.

If, for example, a man wishes to get the news, he may punch six tuning buttons on his set one after the other and listen to the offering of each station just long enough to discern that it is *not* news. Finding news on none of the stations, he may turn off the set. In a strict behavior sense, the man has heard, momentarily, the offerings of six stations. But is he "a listener" to any of their programs? Could he be conceived to be "voting for" any of them by his behavior? Of course not. Not only has he not expressed a "preference for" any of the programs, but he has actually "voted against" them as being of no interest to him at that time. This he did by observing each just long enough to discover that he did not want it, then dialing away and finally turning off the set.

The problem of determining "who is a listener" offers a major obstacle to some methods, but here again the coincidental method has no difficulty. The respondent knows whether or not he considers himself to be a program listener. He is the only one who can have this knowledge. If he views himself as a listener, he says so, and to what program and station he is listening if he knows. The coincidental method qualifies, therefore, as yielding a valid measure of program preferences.

C. DOES THE METHOD EMPLOY A UNIT OF MEASUREMENT WHICH IS CONSTANT UNDER ALL CONDITIONS?

One who is "a listener" listens to a program during a period of time. A listener and a time are both involved in the unit of measurement. Is the listener who listens for five of the thirty minutes a program is on the air, a "program listener"? No. He is only a "program-part listener". But there are many such. In obtaining a unit of measurement of listening which is strictly comparable from one program to another how shall these "program-part listeners" be handled?

Two ways suggest themselves. The first is to assume, as do most recall methods, that all programs have an equal proportion of these "program-part listeners" and they may, therefore, be counted as wholes without creating any incomparability between the resulting ratings for different programs. The only difficulty with this course of action is that the assumption is known to be false. It is known, for example, that program-part listening is much greater in the case of variety programs than it is in the case of dramatic programs. The more split up and disconnected the content of a program is, the greater the program-part listening tends to become. Program-part listening is very high in cases like *Major Bowes* where each short act stands by itself. It tends towards a minimum for programs

like *The Thin Man* where the entertainment and interest depend upon understanding all the elements leading up to the climax. When, therefore, "program-part listeners" are counted as "program listeners", the resultant unit of measurement loses all comparability from one program to another. It stretches for the disconnected variety program and shrinks for the unitary dramatic program.

This measure of "total listeners" which includes both "program" and "program-part listeners" indiscriminately is not without a certain value for some purposes. But for the purposes of obtaining program ratings which are strictly comparable from one program to another, it is highly inaccurate.

The second method of dealing with "program-part listeners" is that inherent in the coincidental method—that of giving the "program-part listener" the exact weight in the total which his listening time demands.

The coincidental method's characteristic of evaluating exactly program-part listening is not widely appreciated, but it is readily understood. In this method, calls are made by each interviewer at the uniform rate. In pure random interviewing, the procedure used within each city, each telephone home has equal opportunity of being called. The chances that any family will be called while it is listening to a given program are directly proportional to the length of time the family listens to that program. That is to say, if a person listens to six min-

utes of a one hour program, he has just one-tenth as much chance of being called while he is listening to the program as would be the case if he listened for the full sixty minutes. Conversely, if there are ten times as many people listening for six minutes as there are listening for sixty minutes, the chances are that there will be one "six minute listener" called for every "hour listener". Automatically, the length of listening is weighted by the coincidental method to yield a result which represents "program listeners". By sampling continuously throughout the program and basing the "rating" on all the calls made while the program is on the air, the coincidental method yields a measure which is the "average audience" throughout the time the program is broadcast.

This figure, "average audience", which results when length of listening time is automatically weighted exactly, is the only measure of audience size which is strictly comparable from one program to another, one place to another or one time to another. The coincidental method, therefore, qualifies as using a unit of measurement which is constant under all conditions.

D. IS THE METHOD EQUALLY APPLICABLE TO ALL POPULATIONS AND TO ALL PROGRAMS?

It is on the score of applicability that the coincidental telephone method is limited. First, it can be applied only to telephone home populations. It will

be shown below that for network program ratings this may not be a serious limitation. It is a limitation, nonetheless, particularly because it does not permit analysis of the behavior of population sub-groups by educational level, economic status and race. Secondly, the method is not equally applicable to all programs broadcast. Coincidental interviewing can be conducted efficiently only between the hours of 8:00 A.M. and 10:30 P.M. No network programs are broadcast prior to 8:00 A.M., but there are a few that are broadcast between 10:30 and 11:00 P.M., E.W.T. These programs cannot be measured coincidentally in the Eastern time zone though they can be in all other time zones. For the methods used to cope with these problems see Chapter IX. Unaided by them the coincidental telephone method does not qualify as equally applicable to all populations and to all programs.

E. IS THE SAMPLE REPRESENTATIVE OF THE POPULATION FROM WHICH IT IS DRAWN?

Ratings which are strictly comparable with each other can be obtained only in areas of "equal network opportunity", as was shown in Chapter II.

Of the large cities in the nation in which the conditions of equal network opportunity might be expected to exist, C. E. Hooper, Inc., has selected thirty-two as fitting the requirements for strict equality. The degree of perfection attained in this

company's selection of conditions of listening which are free from the vitiating influence of differences in coverage is illustrated by the following: In the thirty-two cities selected as the Hooper laboratory, one network has sixteen stations of fifty kilowatts power, fourteen of five kilowatts and two of one kilowatt. The average evening rating from 8:00 to 10:00 P.M. for the three months, August, September and October, 1942, was, for the fifty kilowatt stations 8.4. For the sixteen stations of five kilowatts or less, the figure was 8.2.

Another network has in these thirty-two cities only two stations of fifty kilowatts power. Of the remaining thirty, fifteen are five kilowatts, eight are one kilowatt, one is 750 watts, one is 500 watts and five are 250 watts. The average rating for the seventeen stations of five kilowatts or more on this network for the above period was 4.1, while the corresponding rating for the fifteen stations of one kilowatt or less was 3.8. The slightly lower ratings for the smaller stations of both networks is readily accounted for in terms of smaller expenditures for programming during "non-network sponsored time" as compared with the heavier expenditures and elaborate production of network programs.

No other population lends itself to pure random sampling as readily as does a telephone sample. Every home listed is equally accessible to the interviewer, a condition which does not hold to comparable degree in house-to-house interviewing methods.

The high economic levels may prove quite inaccessible in house-to-house interviewing. In many cases, interviewers are not permitted to enter the better apartment buildings. But the telephone reaches the swank and the swankless alike. However, even with a telephone sample, it is doubtful practice to call homes in alphabetical order. If such a procedure is used in large cities, homes of a single racial or religious group may be called by one interviewer for some hours. This difficulty may readily be overcome by numerous devices for selecting from different parts of the directory numbers to be called consecutively.

Because a telephone sample lends itself to a maximum of random selection, there can be but little question that the sample obtained is highly representative of the population sampled—the telephone population of cities in which all four networks have local outlets, if the measurement obtained is a “network program rating.”

Quite apart from the question of how well this sample represents telephone homes, there is another question of importance to network advertisers. In some cities only about thirty per cent of the families have telephones while in others eighty per cent of the families may have telephones. Since telephone ownership tends to be related to economic status, it is apparent that a telephone sample underweights the low economic levels. It covers the group in which a large proportion of the

buying power is concentrated and by the same token tends to eliminate the group with low buying power.

Sometimes high buying power is not important to the advertiser. This may be the case with an advertiser who is selling a low priced bar soap or one who is chiefly concerned with building public good will toward his company and toward American industry as a whole.

Other network advertisers sometimes desire to know how closely data on network programs obtained from a telephone sample resemble those obtained from a cross-section sample of the same area.

Some of the early experimental work with the mechanical recorder method seemed to indicate that such differences might be large and important. Without taking a position for or against the telephone sample in this regard, it may be pointed out that these early figures were based on very small samples. If a sample of 200 homes were used and divided into two equal groups (telephone and non-telephone homes), each would be 100 homes. A rating of 30.0 (meaning thirty listening homes in either sample of 100) obtained in such a sample would have a statistical tolerance (significant difference) of plus or minus 9.0. This means that unless the difference between the ratings for the telephone homes and that for non-telephone were 9.0 or greater there is no statistically reliable difference between them.

The present writers have studied this problem from a number of points of attack:

1. Station Listening

In a sample of 907 personal interviews the respondents were asked the following questions:

"To what station do you listen most in the daytime?"

"To what station do you listen most in the evening?"

The responses for the cross-section sample are compared with those obtained in the telephone homes of the sample in Table V and Table VI.

TABLE V

Responses to the question: *To what station do you listen most in the daytime?*

Station	Telephone Homes	Cross-section of Homes
WAAA	45.6%	46.3%
WBBB	33.9	33.5
WCCC	17.8	17.5
WDDD	1.2	0.9
Others	1.5	1.8

TABLE VI

Responses to the question: *To what station do you listen most in the evening?*

Station	Telephone Homes	Cross-section of Homes
WAAA	46.0%	45.8%
WBBB	40.0	38.6
WCCC	9.6	11.2
WDDD	2.0	1.7
Others	2.4	2.7

The data shown in these tables indicate that there is no significant difference between a telephone home sample and a cross-section sample in their preferences for network stations.

In another study of station listening made by personal interviews in 1,921 rural homes, the respondents were asked: *What station did you listen to last?* In this case the validity of the responses was checked coincidentally where the set was in operation. Where it was not in operation the response was checked by reading the dial setting. The "last listening" reported for the client's station in rural telephone homes and in the total rural sample are shown in Table VII.

TABLE VII

Responses based on the question: *What station did you listen to last?*

Sample	Last Listening to WLW
Rural Telephone Homes	55.9%
Cross-section of Rural Homes	55.8

These figures are even closer than would be expected from two studies made on the same sample. They show that for the given network station there is no difference in the sample of rural telephone homes and cross-section of rural homes in "Last Listening."

2. Network Program Preferences

In a study of preferences for radio personalities based on 2,133 personal interviews in one locality the respondents were asked: *Who is your favorite radio personality?* The rank order of the top ten rated personalities is shown in Table VIII, where the rank in the telephone homes of the sample is compared with that for the total sample.

TABLE VIII

Responses to question: *Who is your favorite radio personality?*

Program	Telephone Homes	Cross-section of Homes
Major Bowes	2	1
Lowell Thomas	1	2
Local Commentator	4	3
Information Please	3	4
Easy Aces	5	5
Aldrich Family	7	6
Jack Benny	6	7
Kay Kyser	8	8
Lux Radio Theater	9½	9
Kate Smith	9½	10

The close relation between the rank orders indicates that for the well-known programs or program personalities there is no significant rank order difference between the preferences expressed in telephone homes and in a cross-section of homes.

3. Program Preferences

In a study of program listening, a printed roster of thirteen programs was used in personal interviews in 2,076 homes. The respondents were asked to indicate which programs they had listened to within the past week. The results are shown in Table IX.

TABLE IX

Printed Roster Recall Study in Eleven Cities

Program	Recall in Telephone Homes	Recall in Cross-section of Homes
Burns and Allen	44.3%	42.3%
Al Pearce	29.0	27.1
Believe-It-Or-Not	26.8	24.9
Paul Sullivan	25.0	24.3
Bob Crosby	25.5	24.2
Blondie	20.8	24.1
Elmer Davis	25.6	24.0
Hour of Charm	27.0	23.3
Uncle Walter's Doghouse	22.4	22.9
Cavalcade of America	22.6	22.2
Show Boat	24.0	21.9
Uncle Ezra	22.3	21.8
Plantation Party	15.3	14.9

These thirteen programs cover a wide range of appeals but in only two cases is the difference between the figures for the telephone homes and the cross-section sample sufficiently large to be considered statistically significant. The one is *Blondie*, reported more frequently in the cross-section sam-

ple, and the other is *Hour of Charm*, reported more frequently in the telephone sample.

It is to be noted that the above comparisons have been made on network programs and network stations. The character of network broadcasting is, generally speaking, somewhat different from that of local station broadcasting. Few network programs are directed toward specific population groups. Rather their appeals tend to be broad.

In the case of many local stations, on the other hand, programs are designed for specific population groups. One local New York station may broadcast only programs which appeal to the high cultural levels. Another New York local station may direct its programs to the lower cultural levels. Still others broadcast to foreign language groups.

If programs of local stations appealing to the lower cultural levels and to foreign language groups are compared in a telephone sample and a cross-section sample, larger differences are found than is the case when network programs or stations are compared in the two samples. For example, a study based on 7,300 personal coincidental interviews in a mid-western city showed that one of the local stations which broadcast foreign language programs received in the afternoon 3.4 per cent of the listening in telephone homes and 6.2 per cent of the listening in a cross-section of homes. However, in connection with network broadcasting, the evidence indicates that the differences between listening in

telephone homes and in a cross-section of homes are relatively small.

Some of the A. C. Nielsen Company's findings on set-tuning based on the mechanical recorder have a bearing on this problem. Data from the Nielsen Radio Index for April, 1943, which were given general release in the summer of 1943, showed by economic groups the percentage of families in which the radio was turned on during some part of each hour. These figures indicated that no significant differences in amount of tuning time exist between the three economic levels in the evening from 6:00 P.M. to 11:00 P.M.

F. IS THE SAMPLE SUFFICIENTLY LARGE TO YIELD STATISTICALLY SIGNIFICANT FIGURES?

In order to answer this question it is necessary to examine individually each of the uses to which the coincidental method is put and consider each sample size.

1. Program Ratings (Based on Total Homes in Samples)

Coincidental calls are made continuously during the time a program is on the air. It follows that the size of the sample of homes called during any program is typically related directly to the length of time the program is on the air. In the Hooper continuous "National" coincidental operation, the

smallest sample on which a program broadcast in all thirty-two cities is rated is 630, the number of homes called during a quarter-hour program which is broadcast only once a week. The sample for half-hour programs is twice that, or 1,260 homes. That for an hour program being taken during twice as long a period is again double the half-hour size, or 2,520; and for a quarter-hour, five-a-week program, the duration of the interviewing is one and one-quarter hours, bringing the total calls to 3,150.

Ratings which are based on samples of less than 600 homes are labeled "indicative only" and the client is furnished with a chart (see Chapter IV) from which he can determine the maximum expected variations in his figures.

Since there are very few quarter-hour, one-a-week network programs, it is apparent that the average Hooper sample for program ratings is large.

2. Program Recruiting Efficiency

(Rating Based on the Available Audience)

Ratings based on total homes in the sample are required to reveal the proportion of the population attracted to a program. They do not, however, tell the full story of the attraction value of a program. The most important segment in a population of total homes is one which varies from hour to hour, day to day, month to month, geographic area to geographic area, from one size of locality to an-

other and from one economic level to another. This segment is the available audience—those radio homes in which someone is “at home and awake”. Homes in which the family is asleep, at the movies, or shopping are no part of the available audience of any radio program.

The available audience in one segment of a population may be eighty-five per cent of the total homes and simultaneously in another it may be only fifty per cent of the total homes. Before the war, the average available audience in areas of equal network opportunity for Monday evening was 80.3 per cent while that for Sunday evening was only 71.7 per cent of the total homes. A Monday evening and a Sunday evening program may each obtain a rating of 20.0, based on the total home sample. But when the size of the available audience for the two nights is taken into consideration, the Sunday program is seen to have appreciably stronger attraction value.

Similarly, a program may be found to obtain a rating of 20.0 based on total homes in both large cities and small towns. But the attraction value as determined by the percentage of the available audience recruited by the program may show them to be quite different.

In order to measure completely the comparative appeal of a program in sub-groups of a population, ratings based on the available audience are essential. In their absence erroneous inferences and

conclusions are unavoidable. Some of these errors, particularly as they apply to measurement of program appeal in different economic levels, will be considered in later chapters.

Data on the size of the available audience is a fundamental requirement in measuring the attraction values of programs and in comparing the relative appeal of a given program in sub-groups of the population. Nevertheless, those who use methods other than the coincidental, quite overlook or disregard this fundamental. Of the methods currently in wide use, the coincidental is the only one which can measure the available audience during the broadcast of any program. The day-part recall and the printed roster typically employ only samples of homes in which someone is at home *at the time of the interviewer's call*. This sample is quite different from one in which some one was at home *at the time of the broadcast*. The mechanical recorder ratings are based necessarily on radio home samples. The recorder reveals that the set is turned on or turned off. But when it is turned off (and in some cases when it is turned on), the family may be at home or it may be at the movies, on vacation, or fast asleep. The recorder yields no evidence on the size of a program's available audience either in a total population or in population sub-groups. Coincidental ratings based on the available audience are furnished to individual advertisers where the sample is 600 homes or greater.

3. Pattern of Listening

It is important for the advertiser to know not only the overall attraction value of his program, but he also needs to know which parts of his program have strong and which parts weak attraction value. For this purpose he requires data which reveal the internal pattern of listening to his program.

Because coincidental calls are made continuously while the program is on the air, the listening during any part of the program can be determined readily. The pattern of listening may be studied by one-minute, three-minute or by five-minute intervals, or by units corresponding to the airing of specific features of individual programs. Each individual internal rating on the program which is used for this purpose in the Hooper operation is based upon samples of a minimum of 600 homes. The value to the advertiser of knowing the listening pattern of his program is readily apparent. From it he can determine whether or not he is "inheriting" a good audience, how well each successive "act" in his program is holding what he inherits, whether or not his program is an audience builder or an audience dissipater, how well he is holding the audience during his commercials, whether or not his commercials are located in the most advantageous position in the program and how well his various talent attracts or holds an audience.

Chart IV, A shows the pattern of listening to

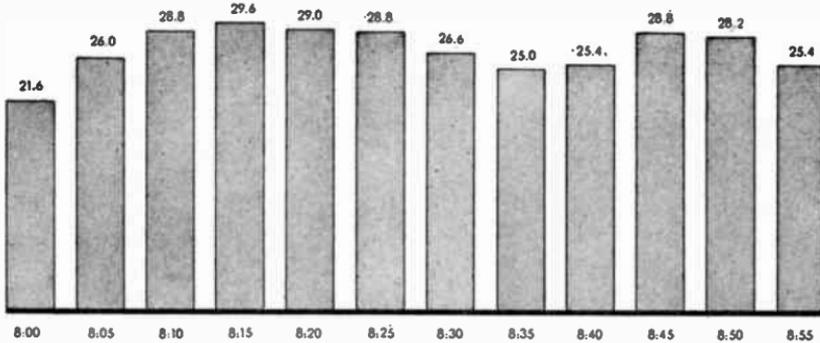
CHART IV

PATTERNS OF LISTENING TO CHARLIE McCARTHY

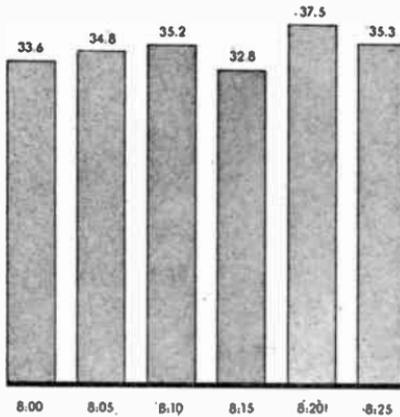
SHOWING INFLUENCE OF CHANGE FROM ONE HOUR TO HALF HOUR IN LENGTH

Rating for each five minutes based on 600 homes

A. One hour in length - October, November, December, 1939



B. One half hour in length - December, 1942, January, 1943



Each rating is the percent of the total homes called during a 5 minute interval which reported listening to the program

Charlie McCarthy when it was a one-hour show. The program was a strong audience builder, building from 21.6 to 29.6 within the first fifteen minutes. The chart reveals marked variation in the pulling power of the different talent. When the Dummy was on during the third, fourth, fifth, sixth, tenth and eleventh segments, the audience reached its peak. It was appreciably lower when some of the other variety features and the guest stars were on.

These findings suggest that a show with a higher average rating might be built by converting the program into a half-hour and eliminating the less strong features. This was done with results shown in Chart IV, B.

The second chart is drawn from data obtained in December, 1942, and January, 1943. It reveals a tight and satisfactory listening pattern. When the show starts, the audience is already assembled. The audience grows slowly throughout the first fifteen minutes. Following the fifteen-minute station break, the audience drops somewhat and rises to its peak in the following five-minute period. In the final period the audience recedes slightly from the peak.

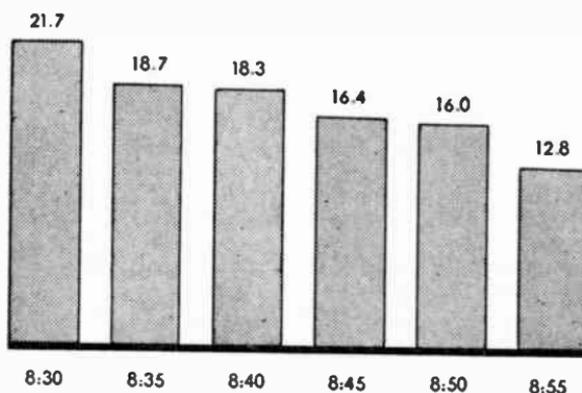
Chart V illustrates the listening pattern of another program which also gets a high rating but is actually an audience dissipater. It inherits a large audience from the preceding program but loses forty-one per cent of the audience while it is on the air.

Each of these three charts indicates that the audience falls off toward the close of the program when most programs have a long commercial. This pat-

CHART V

**PATTERN OF LISTENING
TO AN AUDIENCE DISSIPATING PROGRAM**

Rating for each five minutes based on 600 homes



Each rating is the percent of the total homes called during a 5 minute interval which reported listening to the program

tern is not followed in all programs, as may be seen in Charts VI and VII. Chart VI shows the pattern of listening to *Fibber McGee and Molly* and Chart VII that for the *Bob Hope* program which follows *Fibber* on NBC on Tuesday night.

Fibber is a strong audience builder, building through the first twenty minutes. While there is a

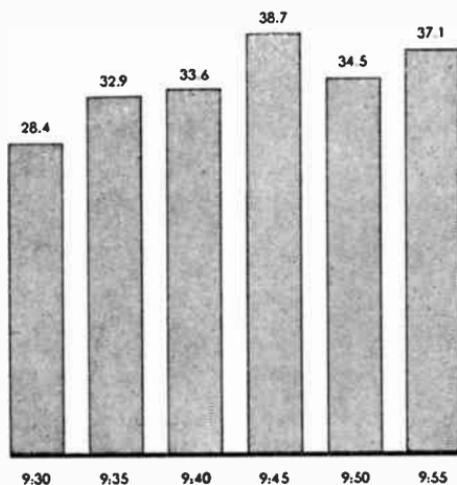
slight dip from the peak listening during the 9:50 to 9:55 P.M. interval, the audience again approaches the peak in the closing five minutes.

CHART VI

PATTERN OF LISTENING TO FIBBER MCGEE AND MOLLY

TUESDAY 9:30 - 10:00 PM, DECEMBER 1942 - JANUARY 1943

Rating for each five minutes based on 600 homes



Each rating is the percent of the total homes called during a 5 minute interval which reported listening to the program

It is possible that this rise in the final five minutes is caused by *Bob Hope*. Chart VII shows that *Hope* comes on with his audience already assembled for him by *Fibber* and that he holds it steadily throughout the half-hour.

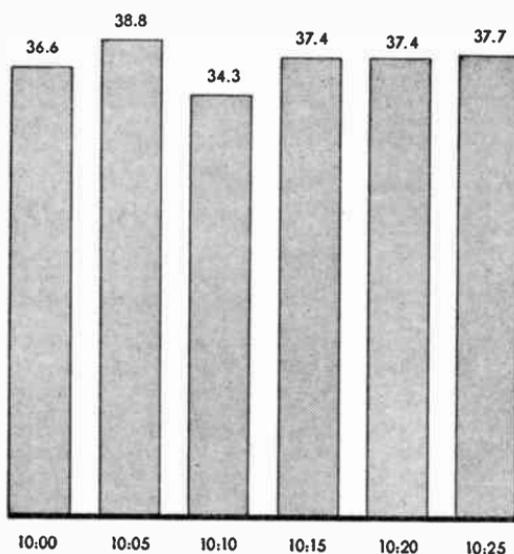
These two charts illustrate an interesting inter-

play of factors influencing program ratings. Both programs are of the same type and appeal to similar audiences. And both are top rating programs.

CHART VII

PATTERN OF LISTENING TO BOB HOPE

TUESDAY 10:00 - 10:30 PM, DECEMBER 1942 - JANUARY 1943

Rating for each five minutes based on 600 homes

Each rating is the percent of the total homes called during a 5 minute interval which reported listening to the program

Hope's rating is helped greatly because *Fibber* builds a large audience which he delivers to *Hope*. On the other hand, *Hope* helps *Fibber*, first by preventing dial changing to catch some program which

follows *Fibber* on another network and also, probably, by causing some tuning to *Fibber* in the final interval in anticipation of listening to *Hope*. Each is a great show made greater because of proper location in a program structure.

4. Measurement of Audience Composition

Network ratings measure the audience size in terms of the per cent of *homes* listening. It is also necessary to determine the number of listeners in each listening home in order to obtain a complete picture of audience size in the given population. Furthermore, many programs are designed to reach certain age or sex groups in a population. No cigar advertiser would select intentionally a program which appeals only to women; nor would a face powder advertiser choose one which appeals largely to men. Knowledge of both the number of listeners in each home and of their age and sex is essential to the efficient use of radio advertising appropriations.

a. Listeners per 100 listening homes. The coincidental method measures the number, age, and sex of listeners per set at the time of the listening, the only time when a program audience exists as such and can be counted family member by family member. Since it is difficult to remember after a broadcast exactly who was listening, any method which does not get these data while a program is being

broadcast must be inferior to the coincidental method in this respect. The degree of inferiority is proportional to the length of time between a broadcast and the subsequent recording of the number, age, and sex of its listeners.

Table X shows the audience composition for individual evenings during June and July, 1943, in

TABLE X

Number, Age and Sex of Listeners per 100 Listening Homes in the Evening

(Based on Listening, June and July, 1943, in Thirty-two Cities of Equal Network Opportunity)

<i>6:00-7:00 P.M.</i>	Sun.	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.
Men	97.4	70.8	65.0	72.0	67.1	68.4	85.8
Women	189.1	118.6	119.8	121.6	125.4	114.7	180.6
Children	41.8	48.6	40.8	47.1	42.8	42.0	88.7
Total	277.8	282.5	225.1	240.7	284.8	220.1	250.1
<i>7:00-8:00 P.M.</i>							
Men	98.8	75.9	74.8	71.1	77.1	71.7	77.9
Women	187.9	117.2	125.8	129.0	120.0	124.4	181.0
Children	48.4	42.9	41.6	41.7	40.8	44.8	88.7
Total	274.6	286.0	242.2	241.8	287.4	240.4	247.6
<i>8:00-9:00 P.M.</i>							
Men	94.6	72.6	81.0	68.8	78.1	74.6	81.1
Women	188.0	123.8	181.2	122.2	181.1	128.7	145.9
Children	40.1	87.9	40.8	48.9	47.4	85.8	42.6
Total	272.7	283.8	252.5	284.9	256.6	288.6	269.6
<i>9:00-10:00 P.M.</i>							
Men	97.8	77.2	84.7	81.5	85.8	80.0	90.4
Women	140.6	182.6	186.0	129.4	186.1	185.8	149.7
Children	88.5	42.7	40.1	88.0	89.4	89.0	87.8
Total	276.9	252.5	260.8	248.9	260.8	254.8	277.4

the telephone home population of the thirty-two large cities of equal network opportunity. Table XI shows for the same months the audience composition in the daytime hours, Monday through Friday combined. Both tables show the number of listeners per 100 listening homes.

As would be expected, the number of listeners per set in operation is appreciably larger in the evening than it is in the daytime. The greatest difference is in the number of men although, generally speaking, more women per listening home are also listening in the evening than during the daytime hours.

Sunday evening enjoys an advantage over other evenings in terms of number of listeners and in terms of adult listeners per 100 listening homes. Saturday is second only to Sunday; and in the period 9:00 P.M. to 10:00 P.M. achieves the maximum for the week. It is interesting to note that, while the adult audience per listening home is also maximum for the week at this time, there are fewer men and more women than during the same hour on Sunday.

These findings are the more important in view of the fact that Saturday evening has sometimes been thought to offer advertisers less opportunity than is offered by other evenings of the week. This belief has arisen from the fact that Sets-in-Use are usually lower on Saturday than, for example, on Friday.

The chief characteristic of the daytime audience,

during the two months covered in Table XI, is the relative constancy of the number of women listeners per 100 listening homes. It increased slightly at

TABLE XI

Listeners per 100 Listening Homes in the Daytime,
Monday—Friday

(Based on Listening, June and July, 1948, in Thirty-two
Cities of Equal Network Opportunity)

Time	Men	Women	Children	Total
8:00- 9:00 A.M.	44.0	119.5	88.2	201.7
9:00-10:00 A.M.	30.5	119.2	89.0	188.7
10:00-11:00 A.M.	28.8	120.8	28.5	172.6
11:00-12:00 A.M.	24.2	119.8	19.9	168.9
12:00- 1:00 P.M.	27.8	124.6	24.2	176.1
1:00- 2:00 P.M.	30.8	128.1	25.9	179.8
2:00- 8:00 P.M.	25.1	120.9	24.0	170.0
8:00- 4:00 P.M.	27.6	118.0	26.6	172.2
4:00- 5:00 P.M.	81.1	115.6	29.0	175.7
5:00- 6:00 P.M.	89.1	117.7	89.7	196.5

mid-day and decreased in the late afternoon, but the variations from hour to hour were small. The number of men listening per 100 listening homes was greatest between 8:00 A.M. and 9:00 A.M., then dropped sharply to rise again at mid-day. It fell from the mid-day level and rose again with the approach of evening. The number of children per 100 listening homes was also highest for the early morning and late afternoon hours.

So-called "children's" programs are offered, for most part, in the late afternoon but, as will be seen on comparing Tables X and XI, the number of children listening per 100 listening homes was

higher in all evening hours than it was in any daytime hour. Of particular note is the large children's audience found between 9:00 P.M. and 10:00 P.M. on Monday, the period in which "Lux Radio Theater" is on the air. The minimum proportion of children in the evening was found between 6:00 P.M. and 7:00 P.M. on Saturday, a period in which the adults per 100 listening homes was relatively high.

Chart VIII shows how programs influence the composition of the listening audience. During the period 5:30-6:00 P.M., Blue Network Company presents children's programs. While children predominate in this audience, there is still a respectable adult audience. The competing CBS and NBC programs are designed for adults, particularly for women; and while the women predominate in their audiences, the men and children also form a respectable group.

b. Listeners per 100 homes in the sample. The above audience composition data are expressed in number of listeners per 100 listening homes. These data are independent of "Sets-in-Use", which also vary from hour to hour and day to day.

The complete picture of number of listeners per 100 homes in the sample (including listening and non-listening homes and homes in which no one is at home) is obtained by multiplying "Sets-in-Use" by the number of listeners per 100 listening homes. Table XII shows the number, age, and sex of radio

listeners in each 100 homes of the sample by hours of the evenings of the week.

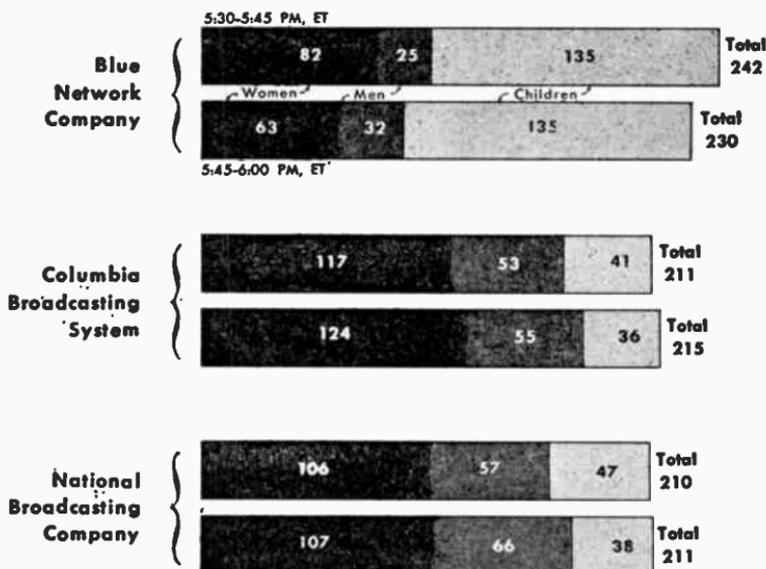
Table XII shows that during June and July, 1943, the number of listeners per 100 homes in the

CHART VIII

AUDIENCE COMPOSITION

MEN, WOMEN AND CHILDREN PER 100 LISTENING HOMES
MONDAY - FRIDAY, JUNE - JULY 1943

Sample size - 1520 listeners



sample was greater on Sunday between 6:00 P.M. and 9:00 P.M. than it was in corresponding periods for other evenings. After 9:00 o'clock, both Monday and Tuesday were superior to Sunday.

The data on Saturday are of particular note. As was remarked above, Saturday evening is widely believed to afford advertisers less opportunity than other evenings. But when both the number of listeners per 100 listening homes and "Sets-in-Use" are taken into account, as they are in Table XII, Sat-

TABLE XII

Number, Age and Sex of Listeners per 100 Homes in the Sample in the Evening

(Based on "Sets-in-Use" and "Listeners per 100 Listening Homes", June and July, 1943, in Thirty-two Cities of Equal Network Opportunity)

<i>6:00-7:00 P.M.</i>	Sun.	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.
Men	22	12	11	12	12	11	13
Women	32	19	19	19	21	19	21
Children	9	8	7	8	7	7	5
Total	63	39	37	39	40	37	39
<i>7:00-8:00 P.M.</i>							
Men	26	16	15	15	16	15	14
Women	39	25	25	27	24	26	24
Children	12	9	8	9	8	9	7
Total	77	50	48	51	48	50	45
<i>8:00-9:00 P.M.</i>							
Men	28	20	22	17	21	19	19
Women	42	34	36	31	35	32	35
Children	12	11	11	11	18	9	10
Total	82	65	69	59	69	60	64
<i>9:00-10:00 P.M.</i>							
Men	30	28	30	25	27	23	24
Women	44	48	49	41	43	40	40
Children	12	16	14	12	12	11	10
Total	86	92	93	78	82	74	74

urday compares very favorably with the weekday evenings. This condition suggests the probability that stronger programming on Saturday evening might make it a superior radio listening evening.

The minimum sample on which program audience composition figures for any program are based in the Hooper coincidental operation is 200 homes found listening to the program. The number of calls made is, of course, many times 200, depending upon the popularity of the program.

5. Program Ratings City-by-City

A program rating based on a composite of the areas of "equal network opportunity" shows the advertiser in terms comparable from program to program, the average achievement of his program in recruiting an audience. But its achievement varies widely from one geographic area to another and from one city to another. For example, it may be found that a program has a "national" rating of 10.0, but, when the rating is broken down by geographic areas, it is 13.0 in one while in another it is only 6.0. When the same program is studied city by city, it may be found that in one city it obtains a rating of 18.0, in another 4.0, and that all of the other thirty cities of equal network opportunity fall in between these two limits.

With such data at hand the advertiser is enabled

to perceive the weak spots in his attempt to attract listeners and is in a position to determine the measures necessary to correct them.

The minimum sample upon which C. E. Hooper, Inc., furnishes individual advertisers with "conclusive" city-by-city ratings is 600 homes. In the published reports for individual cities, all figures based on less than 600 calls are labeled "Indicative Only" and with them is furnished the chart which shows statistical tolerances.

None of the other audience measurement methods currently in use is capable of developing, in the course of its continuous operation, a city-by-city sample of any significance.

6. Sales Effectiveness of Programs

The coincidental interviewer's report contains both the listening data and the telephone number of each home called. This makes it possible to use past records to select continuously a sample of homes which have been found listening to any advertiser's program. This sample of listeners can then be used for the purpose of measuring the sales effectiveness of the advertiser's program. The writers have developed a method for studying the sales effectiveness of programs which is based on samples of "verified" listeners and non-listeners. Through the use of "verified" groups of listeners and non-listeners, the method attains a sharpness of differentia-

tion that is lacking in many forms of "product-use" study.

The verification of listening or non-listening behavior is obtained by calling back homes which have been found to be listening or non-listening in an earlier coincidental interview. On the call-back, the respondent is first asked about his use of products in the class within which the client's brand falls and then is asked if he ever listened to the client's program.

Some of the homes which were found listening to the program in the original coincidental interview, report on the call-back that they never listen to it. These are probably homes in which listening has been very infrequent. The remainder of this group report on the call-back that they do listen. This is the "verified" listener group. They report that they do listen to the program and they had been found listening previously in the coincidental interview.

Similarly for the non-listeners, many of those who were not listening to the program at the time of the original coincidental call report on the call-back that they do listen to the program. They may, of course, have listened at times other than that in which the coincidental call was made. The remainder of this group, those who report that they do not listen to the program, constitute the "verified" group of non-listeners. They report on the call-back that they do not listen and they had not

been found to be listening during the original coincidental call.

This method, then, yields three groups: "verified" listeners, "verified" non-listeners, and non-verified listeners. The most sensitive of all indexes of program effectiveness is obtained by comparing the use of the client's product in the "verified" listener group with that in the "verified" non-listener group. If no reliable difference in product use can be found between these two groups, the sales effectiveness of the program is definitely low.

Chart IX shows the use of the sponsor's brands among "verified" listeners and "verified" non-listeners to eleven selected daytime serial dramas. Some of these programs have been on the air for years advertising old and well-known products. Others have been on the air but a short time and one—Program K—advertises a new brand of product. The difference in the use of the sponsored brand between "verified" listeners and non-listeners to the program is marked in all cases.

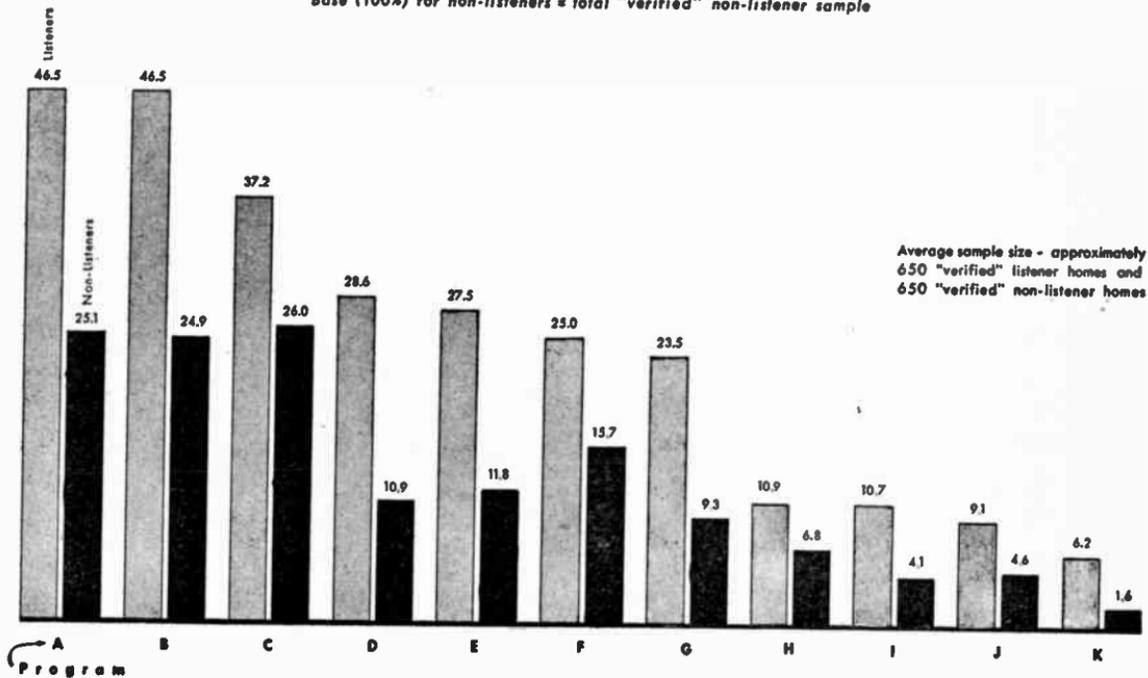
Chart X shows the relation between length of listening in months and use of the sponsor's brand. The sponsor's brand is used by all listening groups more than by non-listeners but, as might be expected, the use of the product increases with length of listening.

The relation between frequency of listening and use of the sponsor's brand is shown in Chart XI. The brand of product advertised by this program

CHART IX
USE OF SPONSOR'S BRAND BY LISTENERS
AND NON-LISTENERS TO ELEVEN DAYTIME SERIAL DRAMAS

Base (100%) for listeners = total "verified" listener sample

Base (100%) for non-listeners = total "verified" non-listener sample



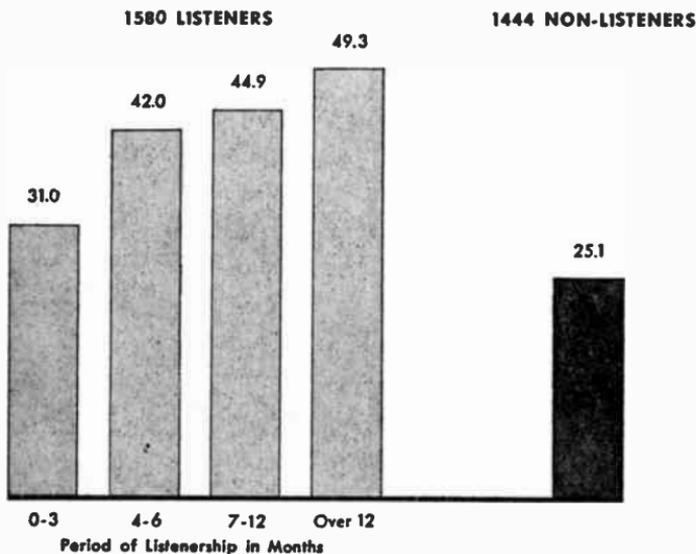
was new. All listening groups use the product more than non-listeners, but the influence of listening four or five times a week is particularly notable.

Charts X and XI show clearly the value of accumulated impressions. The more and more fre-

CHART X

**RELATION OF LENGTH OF LISTENERSHIP TO PROGRAM A
TO USE OF SPONSOR'S BRAND**

Base (100%) is number of listeners who listened for each period of months



quently the sponsor delivers his message to the mind of the listener, the greater is the program's effectiveness.

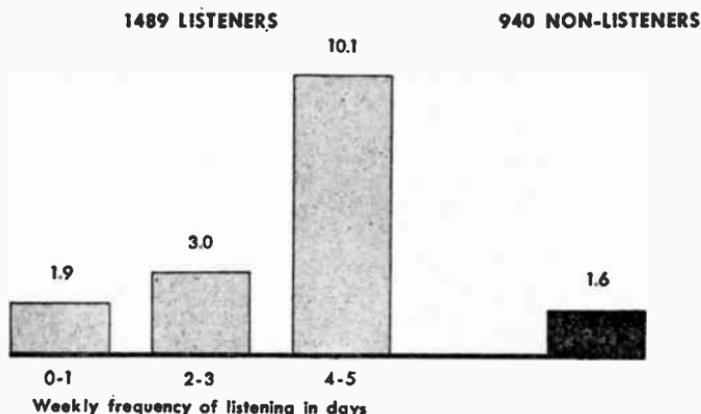
Much misunderstanding of the effectiveness of radio programs exists in the industry. Many adver-

tisers appear to believe that in order to have any sales effectiveness a program must create a "conscious impression" on the listener.

CHART XI

**RELATION OF FREQUENCY OF LISTENING
TO USE OF BRAND SPONSORED BY PROGRAM K**

Base (100%) is number of listeners who listened with each frequency



The importance of the problem is enhanced by the well-known and well-developed human capacity for mentally "tuning out" sounds which are of no immediate interest. The belief that a "conscious impression" is essential to the modification of behavior is based on the assumption that man is a purely rational animal. The belief is, therefore, quite groundless. If it is necessary to produce a "conscious impression" in order to influence a listener, it follows that listeners to a program who are not aware of

the name of the sponsor should use the brand no more than do non-listeners. That no such consciousness is essential to program effectiveness is demonstrated by the results shown in Table XIII.

It will be seen from the table that in the case of each program the use of the product is much greater

TABLE XIII

Sales Effectiveness of Programs in Relation to
"Conscious Impression"

Per Cent of Sample Using Sponsor's Brand

Program	Listeners		Non-listeners
	Identify Sponsor	Do Not Identify Sponsor	
A	8.3	4.3	2.9
B	22.0	19.4	12.2
C	62.4	57.6	47.7

in homes that listen to the program but do not identify the product advertised than it is in non-listening homes.

These results are representative of the conditions found for most programs. In the case of Program B, the use of the product is found to be only a little less among non-identifiers than among identifiers. However, in the experience of the present writers there has been no case where the non-identifiers completely equaled the identifiers in use of the product. It is clear from these findings that a program's effectiveness in influencing the behavior of

listeners is quite independent of any "conscious impression".

While it is true that listeners who identify the sponsor are influenced to a somewhat greater degree than those who do not, it does not follow that the one even contributes to the other. More probably both are the results of something else—the length of listening time. The longer a listener had listened to a given program the greater would be the probability that he would have learned the sponsor's name. By the same token, the greater would have been the program's opportunity to influence the listener's buying behavior. But the important point to be noted is that buying or other behavior may be influenced long before a "conscious impression"¹ is established.

These findings also have bearing on another problem which radio has inherited from the magazine field. It may be called the problem of the "Horse and Cart". The question asked concerning magazines is: Do people who read a magazine come to use a product because of the reading of a client's advertisement, or do they read the advertisement because they already use or are interested in the client's product? Which is the causal factor? Which is the Horse and which the Cart?

The answer to this problem has proved very difficult to obtain in the magazine field. It is not surpris-

¹"Conscious Impression" as reflected in an index to "Sponsor Identification" is obtained only by the coincidental method.

ing that the same question should be raised concerning radio. Specifically the question for radio is: Do people who listen to a program buy the product advertised as a result of the listening or do they listen to the program because they already use the product or are otherwise pleasantly disposed toward the sponsor?

If the program recruited listeners already favorably disposed to the sponsor, two conditions would follow as corollaries. First, all or most of the listeners should identify the sponsor. Secondly, those listeners who do not identify the sponsor should be no more favorably disposed toward the sponsor than are the non-listeners. Both of these corollaries are proved to be false by the results presented in Table XIII. In the case of listeners who do not identify the sponsor, the possibility that the listener listens because he has an interest in the sponsor reduces to an absurdity.

The problem of "Cart and Horse" which has dogged the footsteps of measurers in the magazine field, is solved in large part by the present method of studying program effectiveness. This method has a further advantage in that the use of "verified" groups reveals product effectiveness for new programs long before it can be discovered by most other methods of studying product use. The average sales effectiveness of the program is obtained by comparing the use of the product in the total listening group (verified listeners plus non-verified listen-

ers) with the verified non-listener group. In these over-all comparisons the listening and non-listening groups are weighted according to their distribution in the population sampled.

These studies of program effectiveness are based on a minimum sample of 1,000 previously recorded listening homes and 2,000 to 3,000 homes that have been recorded as non-listening. The size of the latter group depends upon the popularity of the program and is typically larger than the sample of listeners because of the excessive shrinkage in verification.

7. Radio Station Listening

In the early days of radio audience measurements the interest was almost entirely in network program ratings. Radio station popularity was measured largely by the size of the incoming mail bag. However, it soon became apparent that radio listening habits were, to a large extent, local in character rather than national. For example, Denver and Salt Lake City are two areas of equal network opportunity in the Mountain region. However, Station KSL in Salt Lake City may deliver a local rating of 40.0 to "Lux Radio Theater" while Station KLZ in Denver may deliver only a 22.0. Numerous local factors, even in the same geographic area, play an important part in determining the size

of the audience attracted by a program in each locality.

The interest in local measurements has been further facilitated by the realization of radio's power for promoting local business. In earlier days radio was thought to afford competition chiefly to the national magazines. Newspapers were affected, if at all, only as far as national advertisers were concerned.

This situation is changing rapidly. Radio, despite its infancy, is already competing with the newspapers for many kinds of local advertising. Even department store advertising, which newspapers have long held unchallenged, is now courted by radio. It seems probable that radio's vigorous approach will soon develop the techniques required to make even furniture advertisers the same gratifying contribution it has made to soap and automobile manufacturers. In this competition, radio has one great advantage over other media. It is familiar with their techniques and has learned much from them, while they have learned little either about or from radio.

The rapidly expanding local business in radio has given rise to a strong demand for local measurements. To meet this demand, a "Continuing Measurement of Radio Listening" for individual cities was developed in 1940. In this service the local stations of each market studied are furnished with (a) comparative over-all indexes to station listening for individual stations for the morning, afternoon and

evening hours, and (b) program and time-period ratings. The minimum sample on which a Station Listening Index figure is based is 2,400 homes. The minimum sample for a conclusive program or time period rating is 600 homes. Program or time period ratings based on less than 600 homes are labeled "Indicative only". The utility of these studies was demonstrated by the fact that the local stations in fifty large cities adopted them within two years.

8. Listening Area Studies

The cities which constitute the areas of "equal network opportunity" are, of necessity, also the areas of maximum network competition. In all other areas which it reaches, a network program or a station's program encounters less keen radio competition. The large cities are also the areas in which other attractions offer radio a maximum of competition for the attention of the public. As a result of these two factors and some others, the amount of listening to network programs in the smaller network station cities and towns tends to be greater than that in the large network station cities. This is to be taken as a general statement.¹ It does not apply equally to all types of programs.

¹ The Neilsen Index figures which have been given general release would seem to controvert this statement. The winter of 1948 (January) release showed no significant differences between large and small cities, either in the daytime or in the evening, in the per cent of families whose sets are turned on some time during each hour of the day. The spring (April) release showed no significant differences for the daytime but in the evening more tunings in the large city families was indicated.

Radio stations which have a wide coverage area require not only a measure of how well their programs recruit audiences in the equal opportunity area; they need also to know the listening habits throughout the whole area served by their station. For this purpose, a Station Listening Area service based on the coincidental method has been developed. The size of the sample required for these studies, except for certain minimum requirements, depends upon the size of the area covered by the station. In the area study conducted in the WLW listening area in mid-winter, 1943, the sample was 335,000 homes. The smallest sample on which any individual program or time period rating is based in these studies is 400 homes and all figures based on less than 600 homes are labeled "Indicative only".

It is sometimes asserted that the telephone coincidental method cannot be used in either small town or rural studies. The correct statement is that the method cannot be applied as frequently to small town and rural areas as to large cities. The coincidental method may be used in some manner or other wherever and whenever 600 residential telephone numbers are available. In the WLW listening area studies, small towns, villages and rural areas are covered. In the 1942-1943 study, approximately 50,000 homes were interviewed at random in all rural sections of the WLW listening area.¹

¹The results of a control experiment were shown in an earlier section, which revealed that WLW listening in rural telephone homes was identical with that in a cross-section of rural homes.

The variations in a station's audience size in different parts of the area it covers are tremendous. They result from all those factors which constitute variations in competition, station signal strength and habits of listening. Only an intensive study of a station's area will reveal the wide variations in station listening produced by the inter-relations of coverages for all the stations that impinge upon or are included within the service area of the station studied. For example, it may be said that within a certain radius of Cincinnati, fifty per cent of all the listening is to WLW, but within that radius, WLW listening will be found to vary from six per cent of the listening to eighty-five per cent of the listening, depending upon geological and geographical conditions which influence WLW's signal strength, upon the signal strengths and network affiliations of the other stations which compete with WLW for listeners in various parts of its service areas, and upon variations in local interests and habits of tuning.

Reliable area studies require a sample which reflects all of the significant variations in competition, in signal strength, and in listening habits. The sample must be very large. Any attempt to measure the listening conditions in a station's area with a small sample must necessarily lead to absurdity. No matter how "scientifically" one may lay out his sample, the great variations in coverage from locality to locality cannot be reflected in a meaningful manner

by the best possible distribution of a few hundred homes. The differences existing between one part of a station's service area and another are of primary importance. Averaging these differences, as must be done with small samples, hides significant information.

G. SUMMARY

It is apparent from this detailed consideration that the telephone coincidental method meets five of the six fundamental requirements of validity and reliability in a highly satisfactory manner. It yields a valid measure of listening behavior and of program preferences. The unit of measurement remains constant under all conditions. It represents with a high degree of accuracy telephone homes in the areas selected for study. The sample sizes are adequate for each of the purposes for which the method is used. But the method fails to satisfy one fundamental principle—universal applicability. It cannot be applied to all populations; the results cannot be analyzed by educational level, racial, and economic status; and the method is not applicable to programs broadcast before 8:00 A.M. or after 10:30 P.M., local time.

VI. THE COINCIDENTAL METHOD (*Continued*)

BASIC MEASUREMENTS OF THE RADIO INDUSTRY

In the preceding chapter the utility of the coincidental method as it applies to the needs of advertisers and radio station management was considered. In the present chapter, its utility in developing the basic data which reveal the performance of radio as a medium of mass communication will be examined.

Radio is a new medium but radio audience measurement is much newer. No radio research method of wide scope other than the coincidental has a consistent history of operation which goes back, at the present writing, for a period of more than ten months. The Hooper operation, on the other hand, has employed the telephone coincidental technique continuously since 1934. The importance and utility to the industry of the records that have been developed by this operation need no comment. As some erudite historian has remarked, "No one can know

where he is going unless he knows where he has been”.

A. BASIC INDEXES

The four basic measurements of the radio industry—the measurements upon which the condition of the medium is judged—are “Available Audience Index”, “Sets-in-Use Index”, “Sponsored Broadcast Hours Index”, and “Average Rating Index”. By watching closely these four trends it is possible to determine the degree of health enjoyed by the industry as well as some of the factors which have operated to bring about certain of the variations in listening. The value of all such basic trend information lies not in satisfying idle curiosity concerning what has happened to the industry in the past but, rather, in furnishing the means of interpreting current changes and planning for the future. Each of these four indexes for the daytime is based on data obtained Monday through Friday from 9:00 A.M. to 6:00 P.M. The evening indexes are based on data obtained Sunday through Saturday from 6:00 P.M. to 10:30 P.M.

a. *“Available audience index”*. The first of these four basic industry measurements is the available audience size trend. As was remarked above, the available audience is that proportion of the radio homes in a population sampled in which someone is “at home and awake”.

The “Available Audience Index” is the most basic

measurement of all, because it is from the available audience that every radio program must recruit or attract its listeners. The size of the available audience varies within wide limits from one time of day to another, from one day to another, from one geographic area to another at the same time, and from one time of year to another. It is influenced strongly by weather, temperature, pay-days, rhythm of amusement habits, seasonal activities, national and international events.

b. "Sets-in-use index". The American public is composed of busy people. Many duties, activities, interests and many media of mass communication—movies, magazines, books, newspapers as well as radio—are constantly competing for the public's attention.

The second basic industry measurement produced by the coincidental method is that of radio's ability to compete successfully, at any given time, with those other activities and media for the attention of the public. The proportion of the public using its radio sets at a given time, is the measure of radio's success in gaining the public's attention through the combined efforts of all stations reaching a locality. This measurement—the per cent of the total population which is using its radio sets at a given time—is called "Sets-in-Use".

Radio's ability to attract listeners is entirely dependent upon programs, and, since programs change constantly, so do "Sets-in-Use". It changes from

minute to minute, hour to hour, day to day, week to week, month to month, year to year, and from one geographic area to another.

There is a relation between "Sets-in-Use" and "Available Audience", inasmuch as the listening audience must be drawn from those who are at home and awake. The relation is, however, neither one of perfect correlation nor is it constant, otherwise it would be necessary only to measure the one in order to know what the other is. Such is not the case. Both are required.

"Sets-in-Use" is a measure of the per cent of the sample which listened during a given period of time. For example, if the average "Sets-in-Use" for a given Sunday evening is 40.0, that means that, on the average, 40 per cent of the sample was listening during the four and one-half-hour period from 6:00 to 10:30 P.M. In order to express this in terms of "Average Listening Time", it is necessary only to multiply the span of hours by the "Sets-in-Use" percentage to obtain the average number of hours which the population has listened during any given time span. Thus, 40 per cent (listening) x 4.5 hours equals 1.8 hours, the average number of hours of listening in the sample on that Sunday evening. This observation may be expressed in terms of the number of listening minutes per hour by multiplying 40 per cent by 60 minutes, which gives us 24 minutes of listening per hour.

In the same way, the average number of hours or

minutes of listening can be determined for any period of time. Currently, "Average Listening Time" is furnished hour by hour for all days and for each evening. The number of hours of listening varies within the same limits and in the same way that "Sets-in-Use" vary. Because "Sets-in-Use Index" and "Average Listening Time Index" are two expressions of the same condition they are considered to be one index.

c. "*Sponsored broadcast hours index.*" Programs on the air have various origins. Some are sponsored on networks, some sponsored locally, some are network sustainers, others local sustainers. Of all the varieties of origin, those with the largest audiences are the sponsored network programs. Network sponsors put a large amount of money into production, talent and scripts for the sole purpose of building the attraction value of their programs. Radio audiences are built up chiefly by sponsored network programs. This being the case, the third fundamental figure required for judging radio's state of health is a measure of the number of hours per week during which sponsored network programs are broadcast.

The more sponsored network programs that are broadcast in the evening, the greater is radio's chance of competing successfully with other media, at that time, for the public's attention. But by the same token, *the competition between network sponsored programs becomes more keen.*

This measure of the number of hours per week of sponsored network broadcasts (daytime, from 9:00 A.M. to 6:00 P.M., and evening from 6:00 P.M. to 10:30 P.M.) is called the "Sponsored Broadcast Hours Index". The number of sponsored broadcast hours per week varies within wide limits from month to month and from year to year and exerts a powerful influence on the over-all picture of radio and on individual programs.

d. "Average rating index". The above three fundamental measurements pertain to the over-all picture of radio listening. However, people listen not "to the radio" but to radio programs. The proportion of the population sampled which is listening to a specific program is the measure of the program's audience size in the population sampled. The per cent of the homes in the population of the cities of equal network opportunity listening to a specific network program is called the program's "rating".

As in the case of the "Sponsored Broadcast Hours Index", the "Average Rating Index" is based on sponsored network programs only. "Average Rating Index" trends vary in somewhat the same manner as do "Sets-in-Use". But while there is some relation between them, the relation is not perfect. In fact, as will be seen from the accompanying trend charts, a situation may arise in which "Sets-in-Use" show a strong downward trend while the average program rating trend is strongly upward.

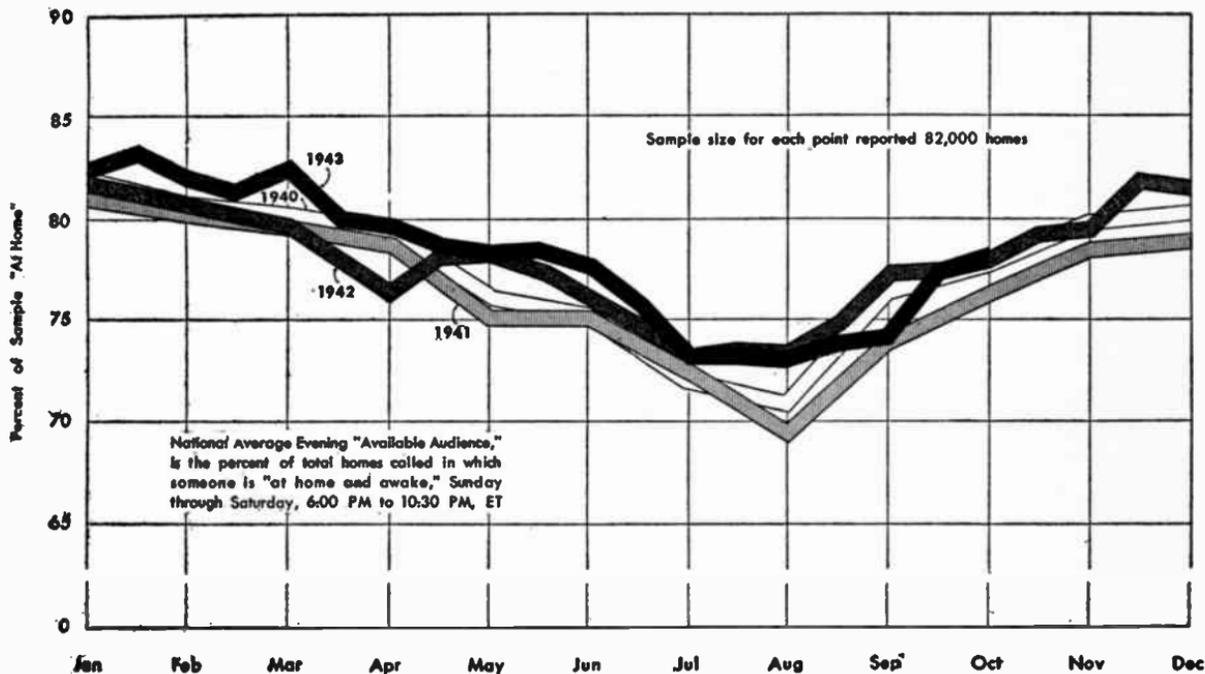
B. THE BASIC TREND LINES—EVENING

The significance of these trends to the industry may be illustrated by a brief examination of the trend charts and some of the factors that they reveal for the war year, 1942. Each of these charts represents the period from January, 1940, to October, 1943, but, since 1942 is the most recent complete year, it will be used for illustrative purposes.

One characteristic of radio which is apparent in all of them is that of seasonal variation. This is much more marked for the evening than for the daytime conditions. Conditions of maximum listening tend to occur in February and those of minimum listening in July and August.

a. "*Available Audience Index*" trend. Chart XII represents the "Available Audience Index" from January, 1940, to October, 1943. In the beginning of 1942, the "Available Audience Index" was at about the same level as in the two preceding years, but in April a major drop occurred which was not found in the two preceding years. In 1940 and 1941, a sharp drop was experienced from April to May, a month later. Daylight Saving Time was initiated late in April, in 1940 and 1941, but War Time was inaugurated in February of 1942. It seems probable that the mild weather in April, 1942, together with the added hour of daylight, accounts for the sharp decline in April evening "Available Audience". If this is the case, War Time advanced

CHART XII
Hooper National Trends
AVAILABLE AUDIENCE INDEX- EVENINGS
 JANUARY 1940 - OCTOBER 1943



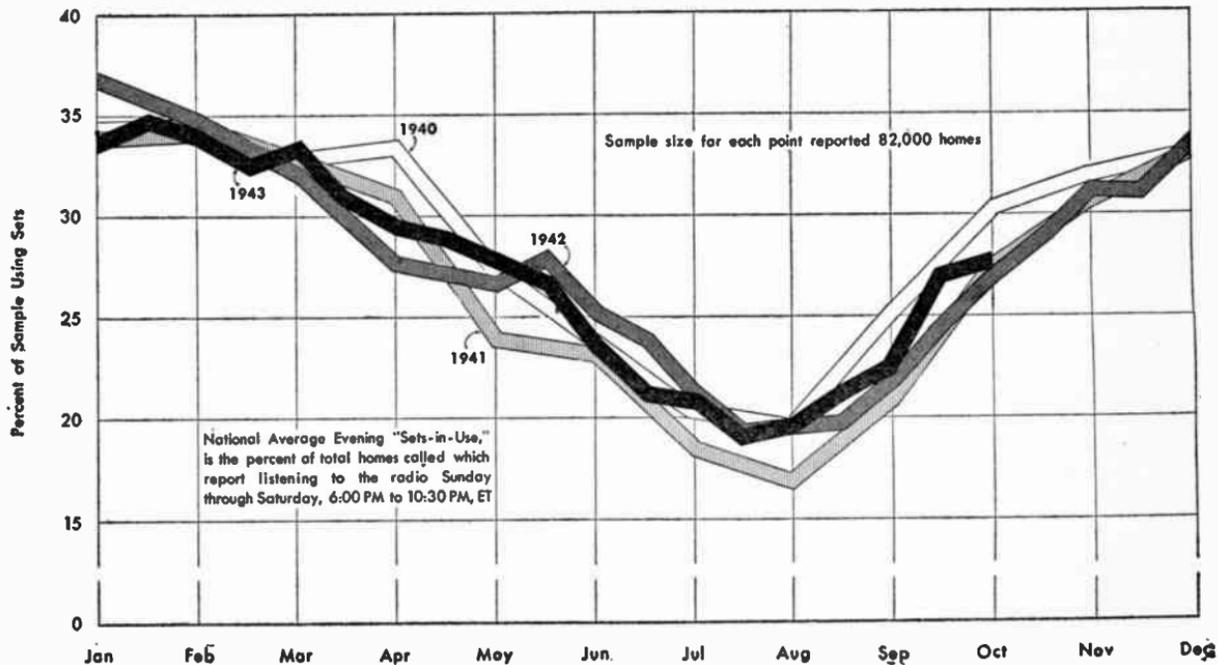
the seasonal decline about a month, and a similar influence should be expected for the duration. (It is also found in 1943.)

In May, 1942, gasoline shortages began to be experienced on the Eastern Seaboard and on May 15th ration cards were issued in that area. At the same time the government requested that travel in other areas be reduced to a minimum. At this time, the "Available Audience Index" reversed sharply the normal seasonal trend. It climbed well above the 1940 and 1941 levels and remained there until the latter part of August. From August through November it was, for the most part, between the 1940 and 1941 levels. In December, following nationwide gasoline rationing, it hit another high for the three-year period. It continued high in 1943.

b. "*Sets-in-Use Index*" trend. Chart XIII represents the evening "Sets-in-Use Index" from January, 1940, to October, 1943. It will be seen that in 1942, "Sets-in-Use" started the year at a high level. This was the month following Pearl Harbor.¹ "Sets-in-Use" remained high in February. In March, 1942, the record was slightly lower than in the two preceding years, but the difference was small. However, in April, 1942, a major drop was recorded which corresponded to the unseasonal drop in the "Available Audience Index".

¹ The influence of Pearl Harbor is not reflected strongly in the December, 1941, figure because December 7th was the last day of December interviewing.

CHART XIII
Hooper National Trends
SETS-IN-USE INDEX- EVENINGS
 JANUARY 1940 - OCTOBER 1943

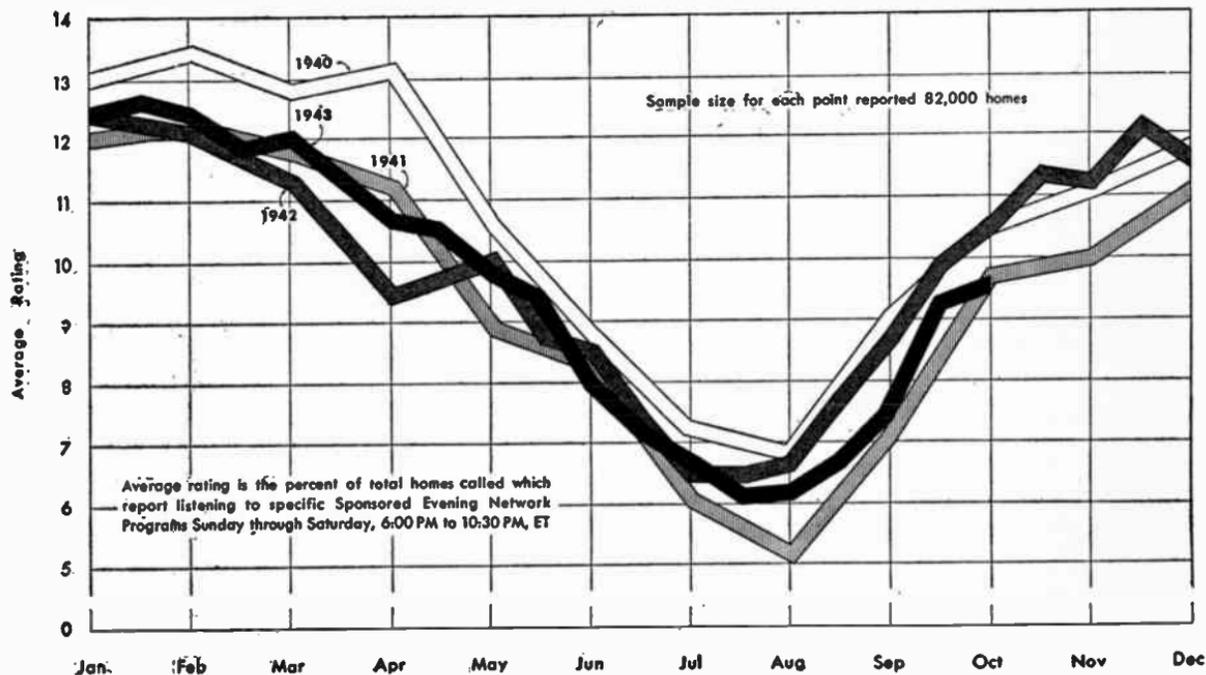


In May, 1942, C. E. Hooper, Inc., adopted the policy of making two evening studies a month instead of one. The first Hooper evening measurement, in May, 1942, was made just before, and the second just after, ration cards were issued in the East. It will be seen from the chart that the second May measurement showed an increase in "Sets-in-Use" instead of the seasonal decline.

"Sets-in-Use" for 1942 remained above the level of 1940 and 1941 until the middle of July. The figures for 1942 remained about at the 1941 levels from August through November. They ended the year following nationwide gasoline rationing in December slightly above the 1940 and 1941 levels.

c. *"Average Rating Index" trends.* Chart XIV shows that one month following Pearl Harbor, program ratings started the year 1942 slightly above the 1941 level but, like the "Sets-in-Use" and "Available Audience" trends, fell off sharply through April. In May, with restrictions in travel, the seasonal trend in evening ratings was reversed and held above the 1941 level throughout the remainder of the year. From the latter part of May through the first part of July ratings were at or above the 1940 level, but from the latter part of July through August they fell between the 1940 and 1941 levels. In November, coincident with the Solomons victory and the British successes at El Alamein, ratings again exceeded the 1940 level—as they did also in the first of December following

CHART XIV
Hooper National Trends
AVERAGE RATING INDEX- EVENINGS
 JANUARY 1940 - OCTOBER 1943



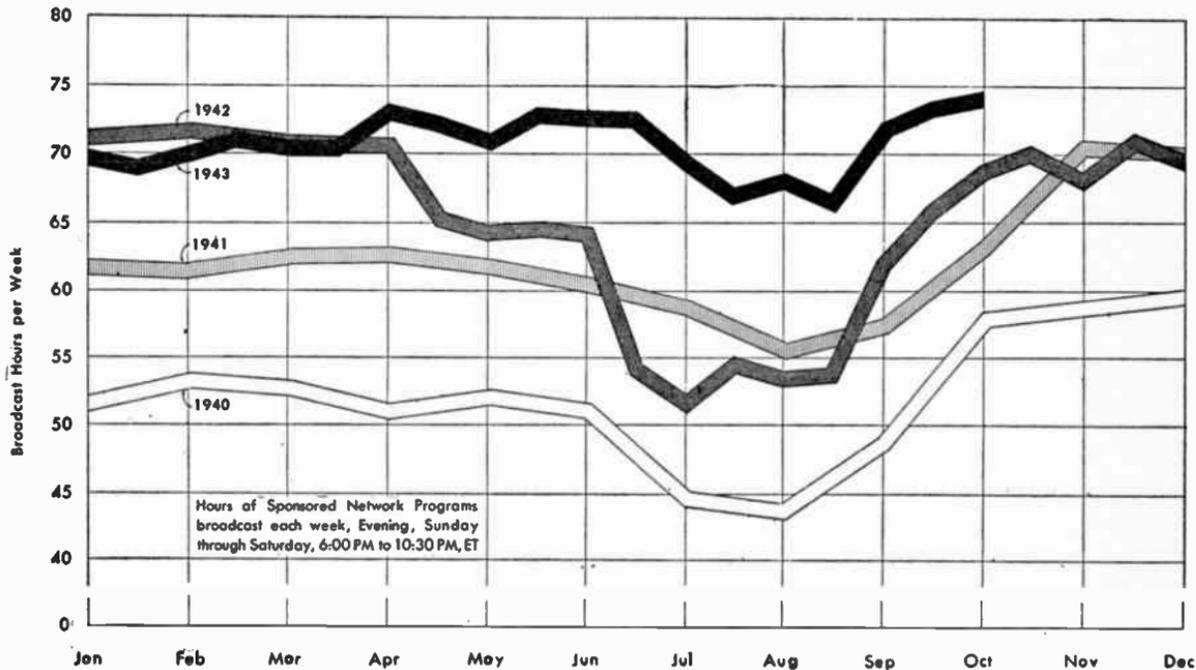
“round three” in the Solomons. Over all, evening ratings in 1942 manifested great strength.

d. *“Broadcast Hours Index” trend.* The strength of the 1942 evening “Average Rating Index” trend is the more remarkable when we examine Chart XV, which represents “Broadcast Hours Index” trend.

Other things being equal, as the number of sponsored network programs increases, the average rating would be expected to decrease because of the keener competition. That is to say, each new network show does not add a completely new group of listeners to those already listening to the radio. Rather, those who are listening tend to be divided up among more programs. The average audience for each program would, therefore, be expected to be smaller. This is the normal relation between the two, the condition to be expected under ordinary circumstances.

But this relation did not hold in 1942. The number of sponsored network broadcast hours in 1942 was greater than in 1941 in eight months of the year and ratings exceeded the 1941 level in nine months. Clearly some new factor operated to enhance evening listening in 1942. This factor was the increase in the size of the available audience which resulted from restriction in travel, an influence which was sufficiently great to more than compensate for those factors such as the growth of the armed forces and increased employment which tend to reduce the size of the available audience.

CHART XV
Hooper National Trends
BROADCAST HOURS INDEX- EVENINGS
 JANUARY 1940 - OCTOBER 1943



C. BASIC TRENDS—DAYTIME

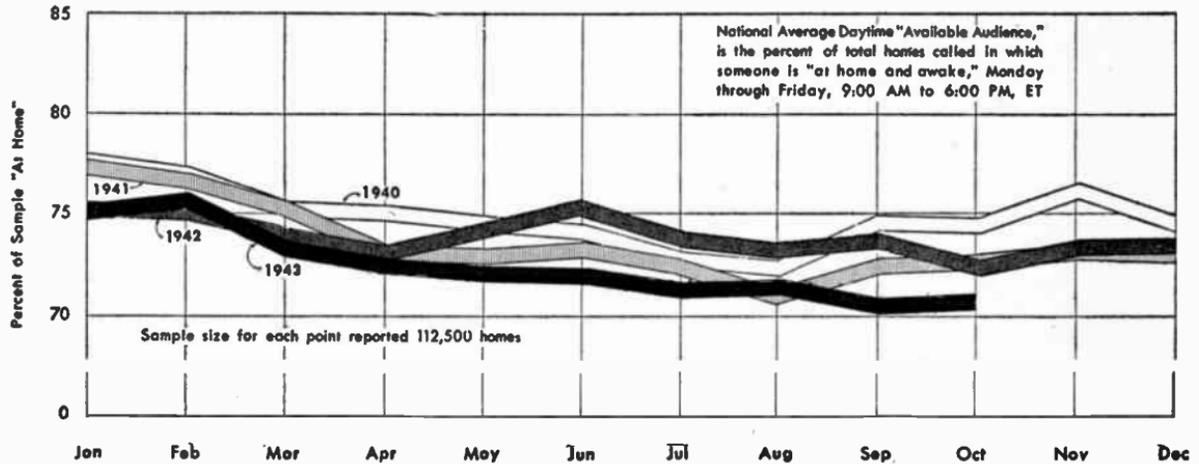
The daytime conditions are represented in Charts XVI through XIX. Each chart is based on Monday through Friday data, 9:00 A.M. to 6:00 P.M.

a. *“Available Audience Index” trend.* Chart XVI shows that the “Available Audience Index” trend for the daytime started the year 1942 well below the 1940 and 1941 levels. This probably reflects greater employment. However, as travel was restricted the daytime audience increased, reaching a three-year high for the month of May. Monthly highs in 1942 were maintained through August and the year closed with the “Available Audience Index” at about the 1941 level.

Two factors—restriction of travel and increased employment operated in opposite directions during 1942. From May through September the restriction in travel more than compensated for increased daytime employment and from October through December, the two balanced each other. However, beginning in December, 1942, gasoline rationing became nationwide, so little or no further gain in available audience could be expected from restriction of travel. Employment, on the other hand, had not reached its peak. It could be predicted on the basis of these data that 1943 available audiences would be lower than those for 1942. The 1943 trend bears out this prediction.

b. *Daytime “Sets-in-Use Index” trends.* Chart

CHART XVI
Hooper National Trends
AVAILABLE AUDIENCE INDEX- DAYTIME
 JANUARY 1940 - OCTOBER 1943



XVII shows the daytime "Sets-in-Use Index" trends. In 1942 this index started the year below the 1941 and 1940 levels and remained below through April. In May, 1942, the seasonal trend was reversed and the daytime "Sets-In-Use Index" rose to a level above that of 1941 but was still appreciably lower than the 1940 level. It remained higher than 1941 through June and in July, 1942, fell to the lowest point reached in the three years, remaining below the 1940 and 1941 levels until November when, following the Solomons victory and the invasion of Africa, daytime "Sets-in-Use" rose slightly above the 1941 figure.

At no time did 1942 daytime "Sets-in-Use" reach the 1940 levels and only in three months, following national and world-shaking events, did it reach the 1941 levels.

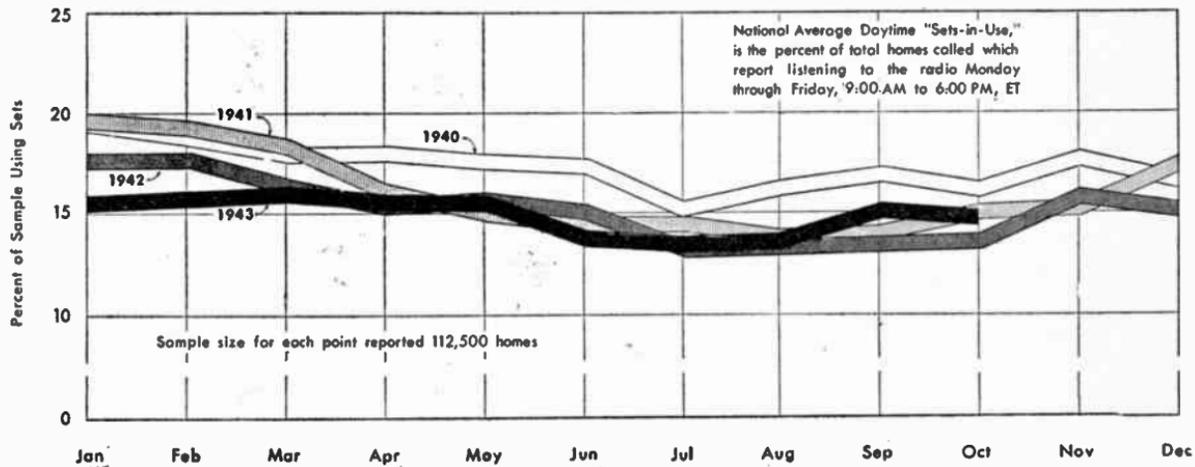
The 1940 level of "Sets-in-Use" was reached in only one month between March of 1941 and December, 1942, a period of twenty months. The exception was in December, 1941, immediately following Pearl Harbor. This is in marked contrast to the evening conditions in which 1942 "Sets-in-Use" made monthly highs for the three years in six of the twelve months. It is also in contrast with the "Available Audience" conditions for the daytime. The "Available Audience Index" fell below the 1941 level in only four months while the "Sets-in-Use Index" fell below the 1941 level in nine months.

CHART XVII

Hooper National Trends

SETS-IN-USE INDEX- DAYTIME

JANUARY 1940 . OCTOBER 1943



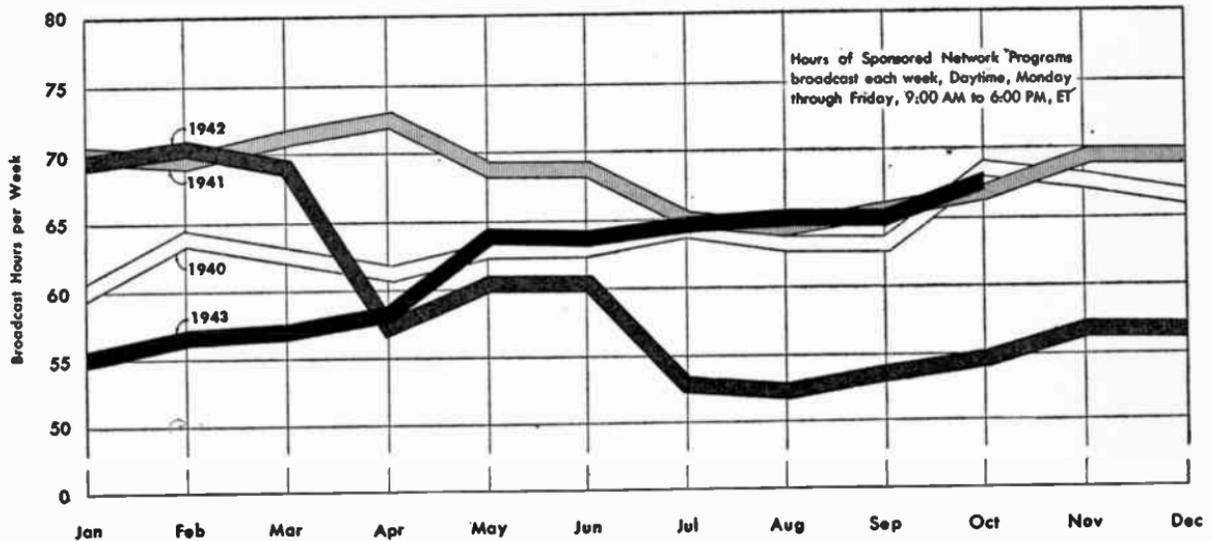
c. "*Sponsored Broadcast Hours Index*" trends. Chart XVIII shows that sponsored network broadcast hours in the daytime increased appreciably from 1940 to 1941, and started the year 1942 at a high level. In April, the "Sponsored Broadcast Hours Index" fell below the levels of the two previous years and remained well below throughout the remainder of the year. The number of daytime network sponsored broadcast hours in December, 1942, was approximately seventy per cent of the 1941 figure. The 1943 line shows a sharp rise.

d. "*Average Rating Index*" trends. The average daytime ratings fell generally from 1940 to 1941 and started the year 1942 lower than in the two previous years. Chart XIX shows that new monthly lows in ratings were made during the first four months of the year; but in May with the restrictions in travel, daytime ratings reversed the seasonal trend and made monthly highs in six of the remaining months.

It is particularly important to note that while average ratings have been relatively high in the daytime in 1942, "Sets-in-Use" have at the same time been low.

The daytime listening situation as revealed in these trend charts is very different from that revealed for the evening. It illustrates well the need and utility of the four basic industry measurements. Consider the facts revealed for 1942.

CHART XVIII
Hooper National Trends
BROADCAST HOURS INDEX - DAYTIME
 JANUARY 1940 - OCTOBER 1943



1. Daytime "Sets-in-Use" was down.
2. Daytime available audience was high.
3. Sponsored network programs recruited larger audiences.
4. There were many fewer sponsored network programs.

From the standpoint of the advertiser a clear-cut conclusion could be drawn. 1942 offered network advertisers an outstanding opportunity to assemble large audiences in the daytime. The available audience was high and the competition weak, with the result that those sponsored network programs which remained on the air paid big audience dividends. The increase in sponsored broadcast hours in 1943 illustrates how advertisers took advantage of this situation.

D. PROGRAM TYPE TRENDS

Similar trend information on program types is also furnished to the industry by the coincidental method. These trends show how the public reacts to each of eight types of programs.

In judging the performance of any type of program, three kinds of trend data are required. The first is the average rating which shows the recruiting value of the type in comparison with other types. Secondly, a measurement is required which shows the share of the listening audience recruited by the programs of the given type. This is called the "Per Cent of Listeners Index". It differs from the rating in that it is calculated as a percentage of

CHART XIX
Hooper National Trends
AVERAGE RATING INDEX-DAYTIME
 JANUARY 1940 - OCTOBER 1943

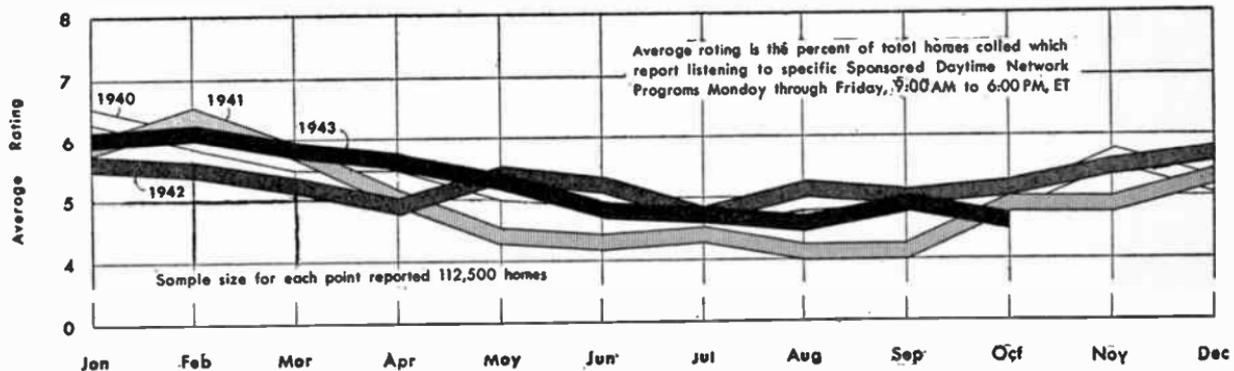


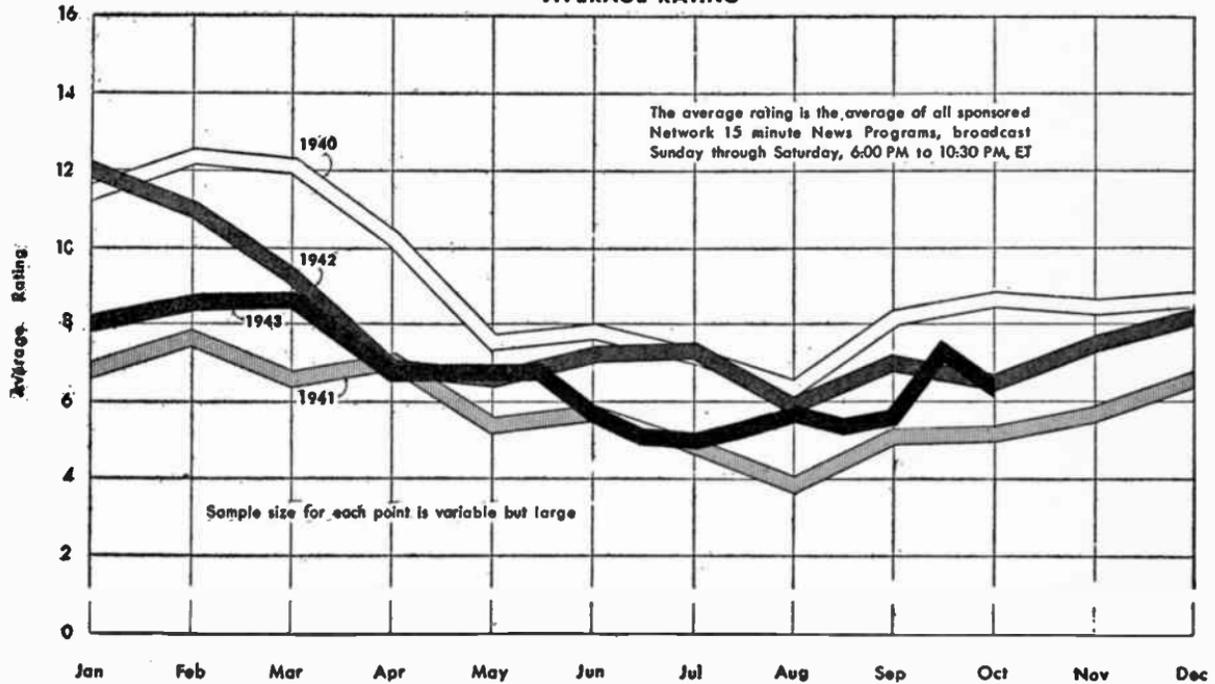
CHART XX

Hooper National Trends

EVENING SPONSORED NETWORK NEWS PROGRAMS

JANUARY 1940 - OCTOBER 1943

AVERAGE RATING



listening homes rather than as a percentage of the total sample. The third index required is that of sponsored network broadcast hours for the type, without which the "Rating Index" and "Per Cent of Listeners Index" trends can have but little meaning. When a new program comes on the air, it does not develop a brand new set of radio listeners. In part, the audience may be new but in part also the program takes audience away from other programs. Beyond a certain point, the more programs of a given type there are on the air, the more ways will the audience interested in that type of program be split up and the smaller will the average ratings for the type tend to become.

The Hooper trends for program types may be illustrated by those for News programs represented in Charts XX to XXII. These trends are based on sponsored network News programs which are fifteen minutes or more in length.

News program trends. Chart XX shows that the 1942 average ratings for News programs started the year at a monthly high. It dropped to the 1941 level in April and remained between the 1940 and 1941 levels for the remainder of the year.

Chart XXI shows the "Sponsored Broadcast Hours" and Chart XXII the "Per Cent of Listeners" trends for News. It will be seen from the latter chart that the average "Per Cent of Listeners" obtained by News programs in 1942 was higher than the 1941 level in eight months and lower in four months.

CHART XXI

Hooper National Trends

EVENING SPONSORED NETWORK NEWS PROGRAMS

JANUARY 1940 - OCTOBER 1943

SPONSORED BROADCAST HOURS

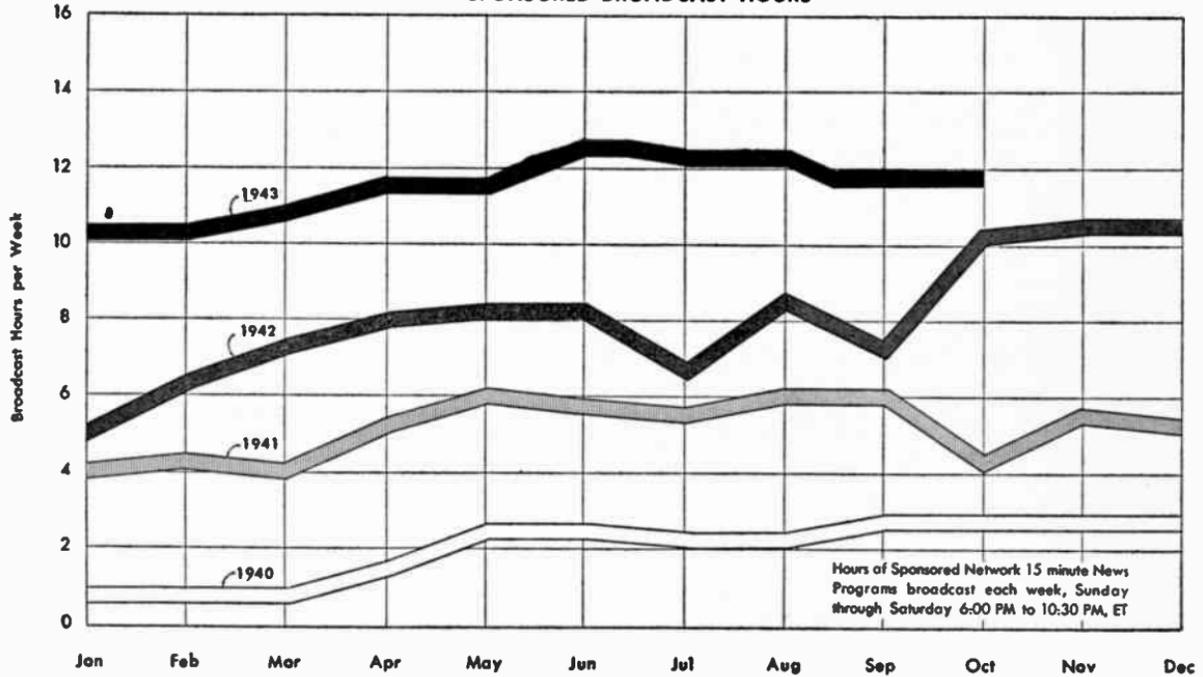


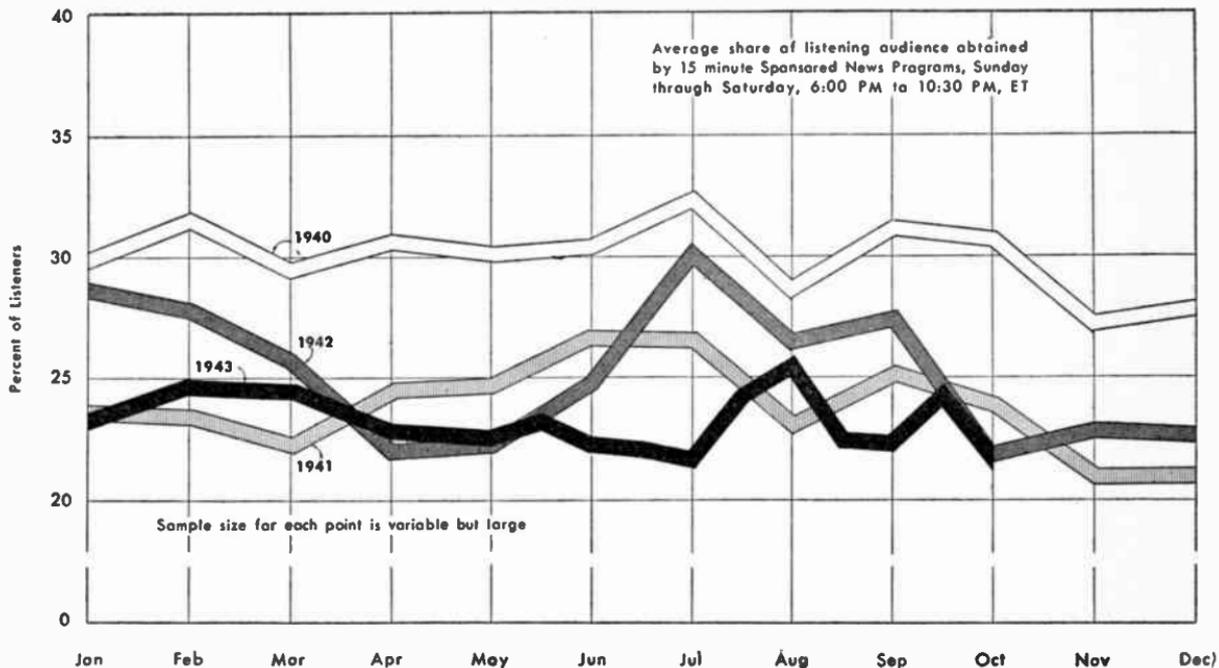
CHART XXII

Hooper National Trends

EVENING SPONSORED NETWORK NEWS PROGRAMS

JANUARY 1940 - OCTOBER 1943

PERCENT OF LISTENERS



The findings for ratings and per cent of listeners do not seem startling in themselves but when they are considered in the light of the trend in sponsored network broadcast hours devoted to News (Chart XXI), it is perceived that the growth of the News program since the beginning of 1940 is one of the more remarkable phenomena in radio broadcasting. Between January, 1940, and December, 1942, "Sponsored Broadcast Hours Index" for News rose over 1000 per cent.

VII. RECALL METHODS

A. DAY-PART RECALL

Recall methods currently play only a minor part in the measurement of radio audiences. Historically, however, they are of primary significance and are still used for some purposes.

The first method that made its appearance in the field was a simple "next day" unaided recall method. In 1929, it was adopted by the Cooperative Analysis of Broadcasting as the standard practice for measuring radio audience characteristics. This non-profit organization was created by the Association of National Advertisers and the American Association of Advertising Agencies. From 1929 to the current writing, the Cooperative Analysis of Broadcasting retained Crossley to produce recall ratings of one variety or another.

Following some experience with recall, the C.A.B. changed to what has come to be known as the day-part recall method. In the early stages of the development of this method, homes were called on the telephone at intervals of four hours throughout the day and asked what programs had been heard in the home during the preceding four hours. Reports

on programs broadcast after 9:00 P.M. were obtained the following morning.

In 1940, C.A.B. shortened the recall interval from four hours to two hours, and in October of 1942 resorted to an overlapping method of interviewing. Following these changes, respondents were asked to report on listening over the preceding two hours, but instead of taking all the interviews covering the given two-hour period at one time, interviews were made at half-hour intervals with the result that some of them were obtained immediately after the broadcast and others one-half, one hour, and one and one-half hours later. This method was used in an attempt to equalize from program to program the length of time elapsing between the broadcast of a program and its recall measurement.

The reliability of day-part recall audience measurements was considered by the present writers in a study¹ made in 1941, designed to determine some of the factors which caused inconsistencies between C.A.B. and Hooper ratings. A second similar study was made at the end of 1941.² These two studies reveal some of the factors which determine the memory values of programs as well as the influence of the shifting base characteristic of recall methods.

Since day-part recall is a memory method, the

¹ Chappell, M. N. *Causes of Inconsistencies Between Day-Part Recall and Coincidental Ratings*, 1941, published by C. E. Hooper, Inc.

² Chappell, M. N. "Factors Influencing Recall of Radio Programs", *Public Opinion Quarterly*, Spring, 1942.

program ratings it yields are influenced by all of the factors which influence the memory for a program. The coincidental method, on the other hand, involves no memory. Comparison of day-part recall and coincidental ratings of the same programs and the analysis of their differences made it possible to discover the influence on memory of numerous factors such as the age of a program, its length, popularity, the use of big name talent, and the influence of elapsed time.

The two studies cited above in which day-part recall and coincidental results were compared were based on two years' ratings and samples of approximately 6,000,000 coincidental calls and the concurrent day-part recall samples for the two years. Only those programs were studied each year for which an average rating for six or seven months obtained by both methods was available.

1. Program Memory Variables

How readily an experience may be recalled depends on numerous factors. Important among them are: the extent and intent of the impression made on the mind by the experience and the length of time which elapses between the experience and its recollection. The influence of the latter factor on day-part recall ratings was recognized early as is attested by the shortening of the recall period from twenty-four to two hours, but the more important variables, those responsible for differences in what

the psychologist would call the degree of original learning—the strength of the impression made—were quite overlooked.

In order to create an impression, work is necessary. Work must be done on the brain through the sense organs. This is as true of radio programs as it is of word lists learned by rote. Further, the more work that is done on the brain by the program the stronger is the impression and the greater the probability that it will be recalled at a subsequent time.

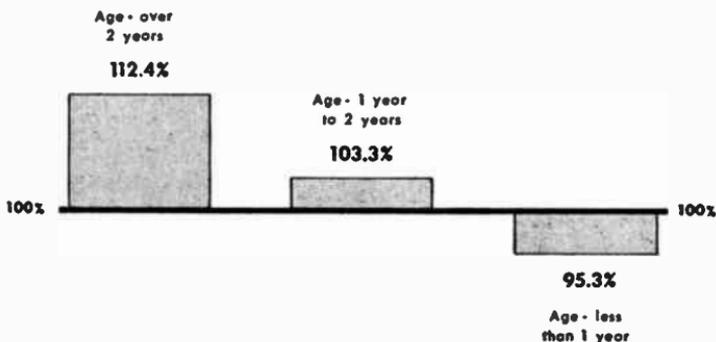
a. Influence of age of program. It would be expected from the above general statements that programs which have been on the air for longer periods of time would be remembered better after any given period of time than would those which have been on a shorter period of time. It should follow, then, that in comparison with their coincidental ratings, the former should obtain day-part recall ratings appreciably higher than those of the latter.

Chart XXIII shows the results of the analysis of the influence of age on day-part recall ratings. Programs over two years of age are remembered about eighteen per cent better than those less than one year of age. These findings are quite in line with the industry's experience. Frequently a new program would go for months with a day-part recall rating of L.T. (less than 1.0) while the coincidental rating of 4.0, 5.0 or 6.0 was reported.

b. Influence of the length of programs. The length of programs would be expected to operate

much as age does. The longer the program is, the more time it has to create an impression on the mind of the listener and the broader is the scope of asso-

CHART XXIII
EFFECT OF PROGRAM AGE

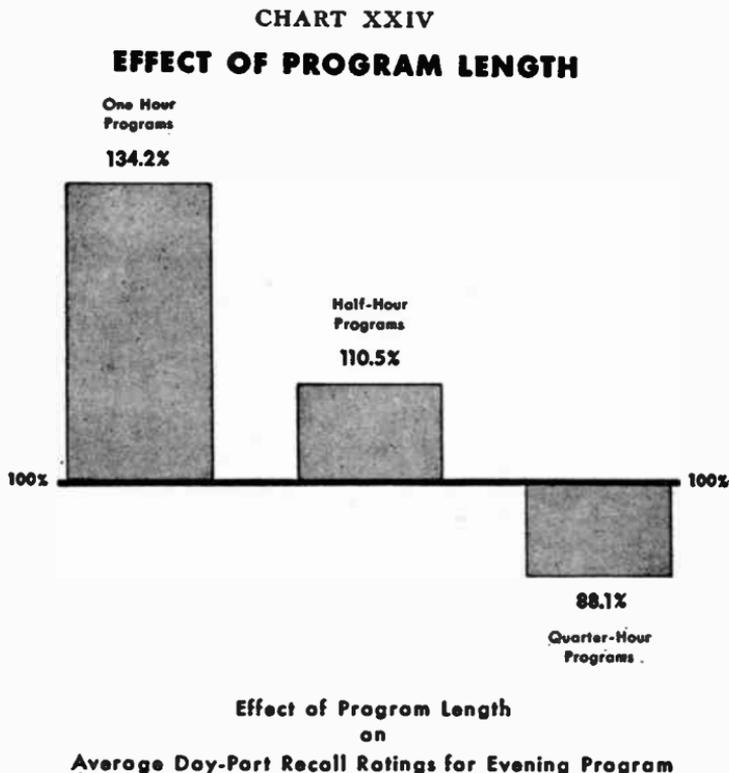


Effect of Program Age
on
Average Day-Part Recall Rating for Evening Program

The average day-part recall rating for programs of different ages are compared with the coincidental rating for those different ages respectively, by expressing the average day-part recall rating as a per cent of the coincidental rating. The findings are based on National Ratings for 106 Evening Sponsored Network Programs.

ciations established. In the language of the psychological laboratory, the longer the practice period, the better the subsequent recall.

The results of a study of programs of the same age but of different length are shown in Chart XXIV. Evening programs one hour in length are remembered about fifty-two per cent better than



The average day-part recall ratings for programs of different lengths are compared with the coincidental ratings for those different lengths respectively, by expressing the average day-part recall rating as a per cent of the coincidental rating. The findings are based on National Ratings for 54 Evening Sponsored Network Programs over 2 years of age.

programs of the same age but only one-quarter hour in length.

c. Influence of program popularity. A program's ability to recruit an audience depends in part upon the breadth and the depth of its appeals. The broader and deeper the appeals are the larger is the audience recruited. That is to say, the very elements in a program which enable it to recruit a large audience should also create, in part, high memory value. The results presented in Chart XXV show that the more popular shows are remembered better than the less. Programs receiving coincidental ratings of 15.0 or more are remembered about 11.0 per cent better than programs which receive coincidental ratings lower than 11.0.

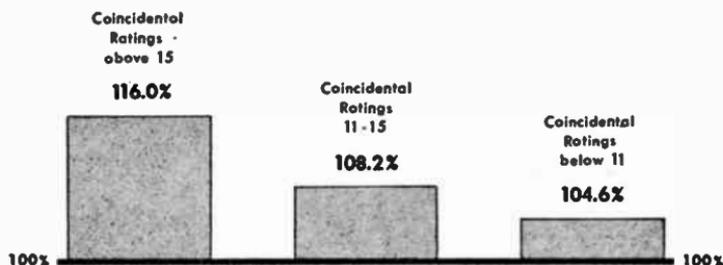
d. Memory values of different types of sponsored network programs. Some further interesting factors in determining the memory for programs are revealed by the analysis of day-part recall and coincidental ratings for programs of different types. The results of such a comparison are presented in Chart XXVI. In interpreting the content of this chart it should be borne in mind that it is based on evening sponsored network programs without regard to length, age or popularity.

Variety programs as a type have the highest memory value. This is understandable in view of the fact that this type is also most popular, that the programs are either half-hour or hour in length and

many have been on the air for years. Much the same may be said of plays.

Perhaps the most surprising fact revealed by the

CHART XXV
EFFECT OF POPULARITY DIFFERENCES



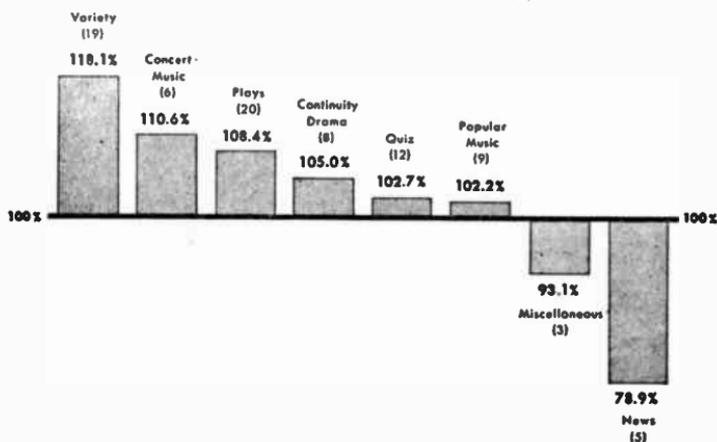
**Effect of Popularity Differences
 on
 Average Day-Part Recall Ratings for Half-Hour Evening Programs**

The average day-part recall rating for programs of each rating category are compared with the coincidental rating for that category by expressing the average day-part recall rating as a per cent of the coincidental rating. The findings are based on 42 Evening Half-Hour Sponsored Network Programs over 2 years of age.

chart is the high memory value for concert music. These programs obtained low ratings as a type but some had been on the air for long periods of time and all were a half-hour or more in length. The

explanation of their high memory value may lie neither in the age nor the length of these programs. It seems more probable that they are remembered well because people who listen to them listen to

CHART XXVI
EFFECT ON PROGRAM TYPES



Effect of Memory Variables
on
Average Day-Part Recall Ratings for Programs of Different Types

The average day-part recall rating for programs of each type are compared with the coincidental rating for that type by expressing the average day-part recall rating as a per cent of the coincidental rating. The findings are based on 7 month's average National Ratings for 82 Evening Sponsored Network Programs.

little else. The shorter the length of a list, the easier it is to remember.

The low memory value of news programs is also noteworthy. Only sponsored network news programs of fifteen minutes in length were included in

this category. None of them, with the exception of Walter Winchell was old. As a type, news programs are medium in rating. But there is probably an added factor which helps explain their low memory value. This factor is the similarity of content of all news programs. Still another factor may be the large number of news broadcasts. Some stations carry a few minutes of news on the hour; others on the half hour. All networks have sustaining, as well as sponsored, news broadcasts. And for all the content is approximately the same.

e. Memory for programs by networks. At the time these analyses were made, the Red Network of the National Broadcasting Company and Columbia Broadcasting System were carrying the older, more popular programs, while the Blue (then an NBC network), with notable exceptions, carried newer, more modest programs and the sponsored evening broadcasts of the Mutual Broadcasting System were predominantly news. The analysis of the influence of memory variables by networks is shown in Chart XXVII. Because of high memory value programs, the average day-part recall ratings for the Red Network and for Columbia were inflated while those for the Blue and Mutual were unduly low.

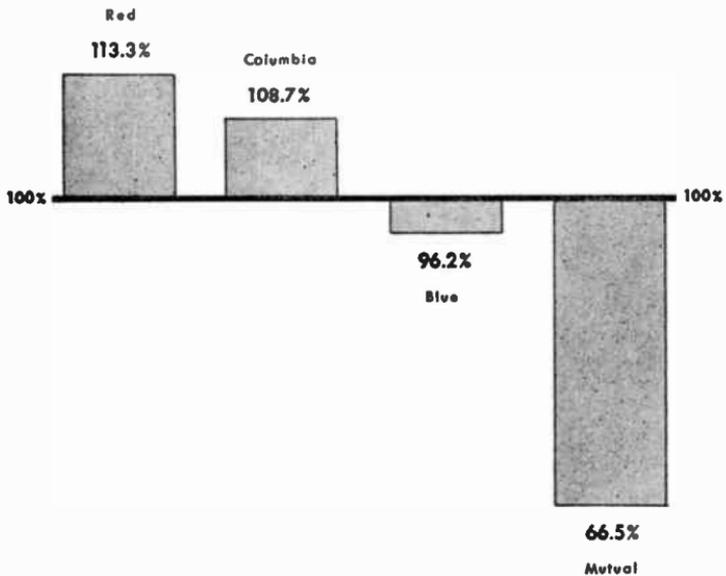
It is clear from these analyses that recall methods do not yield an accurate measure of listening behavior. Recall ratings are limited by all of the factors which limit memory.

2. The Shifting Base of Recall Methods

The second major limitation of recall methods is that they provide nothing which may properly be

CHART XXVII

EFFECT ON NETWORKS



Effect of Memory Variables
on

Average Day-Part Recall Ratings for Evening Programs on Each Network¹

The average day - part recall rating for programs on each network are compared with the coincidental rating for that network by expressing the average day - part recall rating as a per cent of the coincidental rating. The findings are based on 7 month's average National Ratings for 82 Evening Sponsored Network Programs.

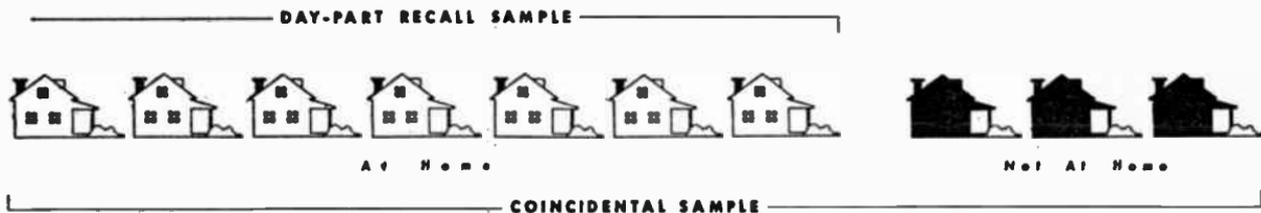
regarded as a base. A true base is a fixed point which remains constant throughout the whole series of events which are to be expressed in terms of it. Thus Polaris furnishes the base for determining directions on Earth because no matter how Earth twists and turns, Polaris is always up there in the same relative position.

The difference between the samples used in coincidental and in recall measurements is shown in Chart XXVIII. Coincidental measurements are based on total homes called, including those in which no one is at home. The base is therefore constant. Recall measurements, however, are based on only those homes in which someone is at home at the time of the interviewer's call. The "not at home" segment of the population is omitted from the base entirely. Such a procedure would offer no major obstacle to comparative measurements if the "not at home" segment remained a constant from time to time and place to place. However, it does not remain a constant. Rather it varies from naught to as much as forty per cent between specific time periods.

These variations in the "not at home" segment of the population in some cases are simple and easily perceived. Such is the case with the seasonal and the geographic variations.

a. Influence of seasonal variation in "not at home". Under normal conditions there is no one at home in approximately fifty per cent more homes

CHART XXVIII
DISTRIBUTION OF "AT HOME" AND "NOT AT HOME"
IN THE TELEPHONE HOME POPULATION



in July and August than in January and February. This condition is presented in Chart XXIX, which shows that out of each ten homes someone is at home in eight and no one is at home in two in the winter, and in the summer someone is at home in seven and no one is at home in three. This variation in the size of the "not at home" segment has a definite influence on recall ratings.

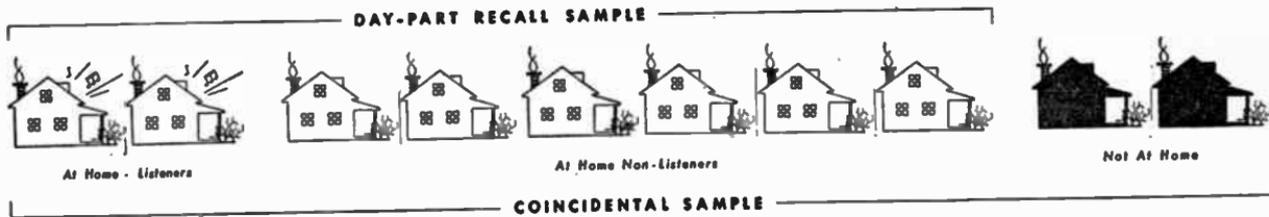
Suppose, as is indicated in the chart, that two of the ten homes were listening in both the summer and winter. The program's coincidental rating, based on the total sample including the "not at home" segment, would be two divided by ten or 20 per cent in both summer and winter. The recall rating, however, is based not on the total homes but only on those in which someone is at home. Under conditions in which two of the total ten homes report having listened, the recall rating in the winter would be two divided by eight or 25 per cent. In the summer, it would be two divided by seven or 28.6 per cent.

Chart XXX represents the findings obtained in 1941 in an actual comparison of day-part recall ratings (based on "at home" sample) and coincidental ratings (based on "total homes called"). The chart shows the following:

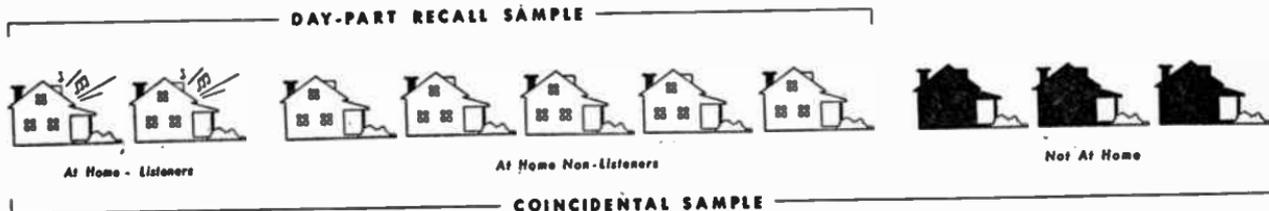
- a. During the winter months, the average day-part recall rating is 9.2 per cent higher than the average coincidental rating.
- b. During the summer months, the average day-part recall

CHART XXIX
DISTRIBUTION OF RESPONSES
IN THE TELEPHONE SAMPLE IN DIFFERENT SEASONS

A. IN WINTER



B. IN SUMMER

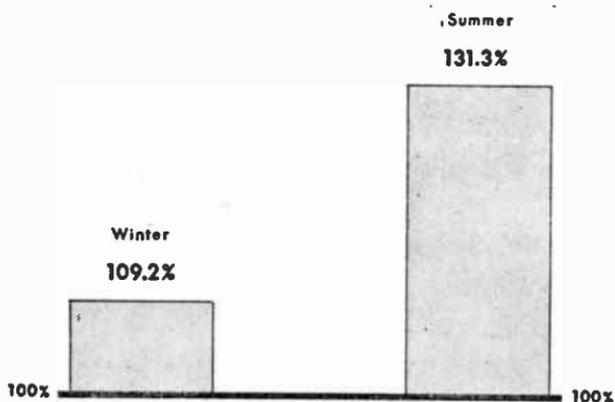


rating on the same programs is 31.8 per cent higher than their average coincidental rating.

As the "not at home" segment increases in size, the recall ratings are inflated to a progressively greater degree.

b. *Influence of geographic differences in "not at home"*. "Not at home" also varies widely from one

CHART XXX
EFFECT OF SEASONS



Effect of Seasonal Variation in "Not at Home"
on
Average Day-Part Recall Ratings for Evening Programs

The average day-part recall rating for programs in January - February and July - August are compared with the coincidental rating for January - February and July - August respectively by expressing the average day-part recall rating as a per cent of the coincidental rating. The findings are based on National Ratings for 46 programs broadcast in both January - February and July - August, 1941.

geographic section of the country to another. During the months, January through July, 1941, the number of homes in which no one was at home in the evening was thirty-four per cent greater on the Pacific Coast than it was in the East.

The normal conditions found in the East and on the Pacific Coast during the summer when "not at home" is maximum, are indicated approximately in Chart XXXI. Out of each ten homes in the East, no one is at home in three; while on the Pacific Coast, no one is at home in four out of each ten homes.

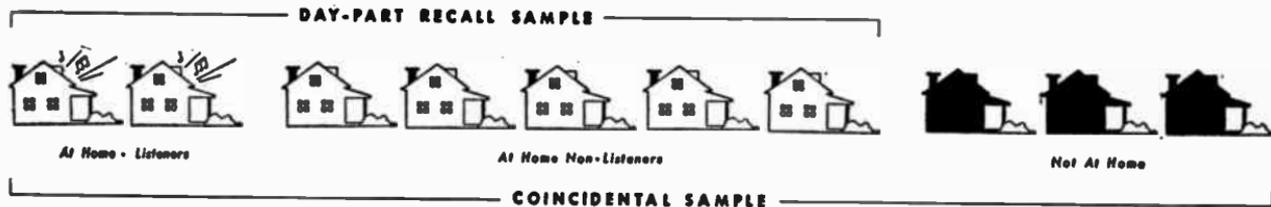
Assume, as is indicated in the diagram, that two of the ten homes report having listened to a given program in each of the two geographic areas. The rating based on total homes, including "not at home", would be two (the number listening) divided by ten (the total homes) or twenty per cent. It would be the same for both areas.

Since "not at home" is omitted in the calculation of the recall rating, only those homes in which someone is at home are used. The recall rating for the East would be two (the number that listened) divided by seven (the number in which someone was at home) or 28.6 per cent; while that for the Pacific Coast would be two (the number that listened) divided by six (the number in which someone was at home) or 33.3 per cent.

Comparison of results obtained on the same programs in the East and on the Pacific Coast by the

CHART XXXI
DISTRIBUTION OF RESPONSES
IN THE TELEPHONE SAMPLE IN DIFFERENT GEOGRAPHIC AREA

A. IN THE EAST



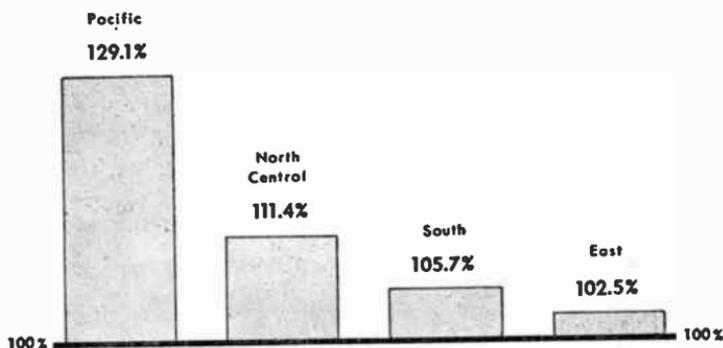
B. ON THE PACIFIC COAST



day-part recall (at home sample) and the coincidental method (total homes sample) are presented in Chart XXXII, which shows the following:

- a. On the Pacific Coast, day-part recall ratings for transcontinental programs average 29.1 per cent higher than their coincidental rating.
- b. In the East, day-part recall ratings for transcontinental programs average 2.5 per cent higher than their coincidental rating.

CHART XXXII
EFFECT OF GEOGRAPHY



Effect of Geographic Variations in "Not at Home"
on
Average Day-Part Recall Sectional Ratings for Evening Programs

The average day-part recall ratings for programs in each section are compared with the coincidental rating for that section by expressing the average day-part recall rating as a per cent of the coincidental rating. The findings are based on 7 month's average sectional ratings for 82 Evening Sponsored Network Programs.

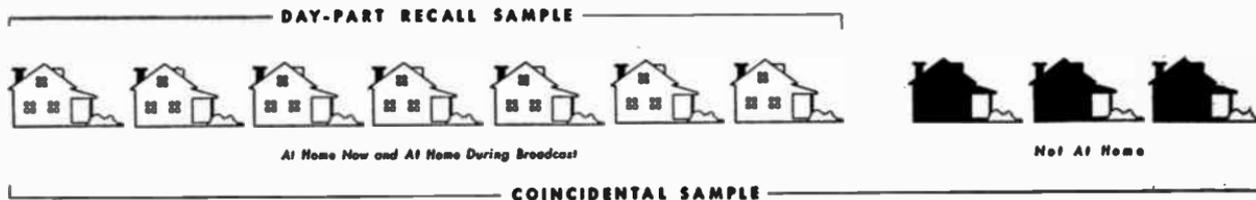
c. Influence of variation in "not at home then". It was remarked above that some of the "not at home" variations influencing recall measurements are readily understood. Such is the case with the seasonal and geographic differences. Variation in the "at home now, but not at home then" segment within any given period of hours covered by a recall study provides another major source of variation and one which is more difficult to perceive.

Suppose that a recall interviewer asks the respondent about programs that have been heard during the two hours just preceding the call. Some of these programs were on the air two hours earlier and some only five or ten minutes earlier.

It is highly probable that anyone who is "at home now" was also at home five or ten minutes earlier; that is to say, the "at home now, but not at home then" segment of the recall sample approaches zero for programs broadcast just prior to the interviewer's call. The situation, illustrated in Chart XXXIII, is very different for programs broadcast two hours earlier. As much as twenty per cent of the sample which is "at home" at the time of the recall interviewer's call may have been "not at home then" (two hours earlier). It follows, then, that the recall ratings for programs broadcast immediately before the interviewer's call will be subject to maximum inflation for they will contain in the base for their calculation no "not at home" of any kind. Those broadcast during the most remote part of the

CHART XXXIII
VARIATION OF "NOT AT HOME THEN"
IN THE DAY-PART RECALL SAMPLE

A. PROGRAM CHECKED IMMEDIATELY AFTER BROADCAST



B. PROGRAM CHECKED TWO HOURS AFTER BROADCAST



recall period (two hours earlier in the above illustration) will receive recall ratings which contain minimum inflation in that the base on which they are calculated will contain a maximum of "at home now, but not at home then".

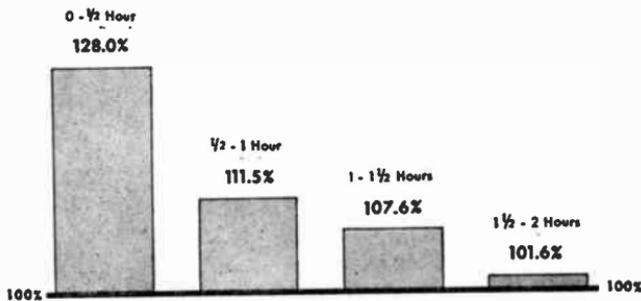
While the "at home now, but not at home then" segment is maximum under the latter condition, it never reaches the magnitude of the actual "not at home" segment. In many homes, no one is at home throughout the day. In others the family is away from home for periods of days or weeks. Neither of these "not at home" groups can possibly be included in the recall sample of "at home now" homes.

The influence of elapsed time on day-part recall ratings obtained in the second 1941 study is shown in Chart XXXIV. It is to be noted that the chart represents changes with elapsed time and not the influence of change in "at home now, but not at home then" only. There is a second variable which operates with elapsed time; this is also in the deflationary direction. This variable is "forgetting". Inflation resulting from the absence of a "not at home then" segment in the base is maximum for programs broadcast zero to one-half hours prior to the interviewer's call. Retention is also maximum for these programs.

The "not at home then" segment increases progressively with elapsed time decreasing the inflation, and forgetting increases with elapsed time further deflating the day-part recall ratings.

Analysis indicates that this total influence of elapsed time is composed of approximately one-third forgetting and two-thirds change in "at home now, but not at home then" segment of the sample.¹

CHART XXXIV
'EFFECT OF TIME LAPSE



Effect of Variations in Elapsed Time Between Broadcast and Checking Period
 on
 Average Day-Part Recall Sectional Ratings for Evening Programs

The average day-part recall rating for programs checked at various intervals after the broadcast are compared with the coincidental rating of the same programs by expressing the average day-part recall rating as a per cent of the coincidental rating. The findings are based on sectional ratings for 82 Evening Sponsored Network Programs.

3. Representativeness of "At Home Now" Samples

From the above considerations it appears necessary to conclude that no recall sample can ever be representative of any population. Variations in

¹ See Appendix of *Radio Program Ratings*, by Chappell, published by C. E. Hooper, Inc., 1941.

“not at home” from one geographic area to another and from one time to another, together with variations in the “at home now, but not at home then” segment inherent in the recall methods, destroy all comparability of recall ratings from program to program, place to place and time to time.

It should be noted that the above findings on the influence of variation in “not at home” may have significance for sampling studies other than those measuring radio audience behavior. Many marketing studies are based on “at home now” samples and make the implicit assumption that the buying or other behavior studied is similar in homes where no one is “at home now”.

Such an assumption may or may not be justified, depending upon the behavior studied. For example, if one wished to study preferences for beer in cans drunk away from home, or to discover the per cent of families using a weed-killer, the assumption would be highly questionable. In the study of fleeting events such as radio programs, the assumption is completely devoid of validity. Its potential effect on the results should be considered carefully in all sampling studies which employ “at home now” samples.

4. Variations in the Recall Unit of Measurement

Recall methods are subject to one further limitation. The unit of measurement is elastic rather than

rigid. In the recall methods, the respondent is asked what programs he listened to during a given period of hours. He may have listened to a program for one minute, five minutes or throughout the whole broadcast. Regardless of the length of time the respondent may have listened, he is counted as a listener if he reports that he was a listener. No accurate distinction can be made in this method between "program listeners" and "program-part listeners".

If the relation between full program listening and program-part listening were constant under all conditions this would not be a serious defection. However, the relation is not a constant one as has been demonstrated by Archibald Crossley in his studies of the mechanical recorder method. In one such study Crossley found that the per cent of tunings in which the set is tuned to any station for periods of eleven minutes or less reaches about seventy per cent during the early part of the evening and about thirteen per cent later in the evening. Furthermore, as was remarked in an earlier chapter, program-part listening varies with program content, being maximum for programs whose content is disconnected and minimum for programs building to a climax, such as plays.

Since the recall unit of measurement varies from program to program and from one time of day to another, the method cannot yield program ratings which are comparable with each other.

5. Conclusion

It is apparent from these considerations that the day-part recall method fails to satisfy most of the basic considerations of validity and accuracy. Because of the influence of memory and the shifting base, it does not yield a valid measure of either listening behavior or program preferences. Its unit of measure which includes both "program" and "program-part" listeners varies from program to program. The random sample employed represents only telephone homes in which someone is at home at the time of the interviewer's call. It does not represent all telephone homes in the areas studied. Finally, while the method is applicable to all programs, an advantage over the coincidental method, it also employs a telephone sample and the results may not be analyzed by economic status, educational level or race.

B. THE PRINTED ROSTER

The printed roster method has never been employed in a national service for reporting on program audiences. However, anyone who has done research in the radio audience field has at one time or another employed the roster method and, in fact, many organizations have had cruel experience as a result of placing too great reliance on figures obtained through some form of the roster method.

Certain characteristics of the roster method that make it highly attractive are: first, it is a personal interviewing technique which, it has been supposed, permits the sampling of almost any population on which data may be required; secondly, through the use of the roster, information on a wide range of programs may be obtained in a single interview.

The essence of the printed roster method lies in the fact that the interviewer presents to the respondent a printed list of programs and asks which of them has been heard within a specified period of time.

The results are necessarily subject to all of the limitations that apply to the day-part recall method, both with regard to the memory variables in programs and the shifting base. There are, however, certain other factors which operate in the printed roster that do not operate in the day-part recall method.

First, there are differences in the size of the "not at home" segments of each economic level. These from one time to another produce serious distortions in the distribution of any sample which is designed to represent a cross-section of all economic groups accurately. The distortion may become so great under some circumstances as to make the resulting data valueless. For example, the number of families "not at home" is greater in the summer than in the winter, but the difference is not equally distributed among all economic groups. Normally

the greatest increase in the "not at home" segment of the population in the summer is found in the A and upper B homes—those people who can afford the resorts or travel.

Suppose that a program received in a roster study made in the summertime, a rating of 10.0 in the A economic group and that only sixty per cent of the A families were at home at that time. And suppose that another program received a rating of 8.0 in the A group in a winter month when eighty per cent of the A families were at home. Which program was heard in more A homes?

When the total A group homes are considered, it becomes apparent that the program with a roster rating of 10.0 in the summer reached 6.0 per cent of the A homes while the program that obtained a roster rating of 8.0 in the winter reached 6.4 per cent of the A homes. These considerations serve to reaffirm the importance of the statement made in an earlier chapter. Only when ratings may be expressed both in terms of the available audience, and in terms of total homes, do they achieve their full analytical significance and prevent serious distortion in implications.

Another element giving rise to variation in roster results is the amount of aid the respondent is given in recalling the program. The more aid the respondent gets the larger will the roster ratings become. For example, if 100 persons are asked, "Do you ever listen to *Spotlight Bands*?" a certain per

cent will say "Yes". More aid is given by asking, "Do you ever listen to *Spotlight Bands* sponsored by Coca-Cola?" and the per cent who say "Yes" becomes larger. If the question is, "Do you ever listen to *Spotlight Bands* sponsored by Coca-Cola on WJZ?" the "Yes" response will be further increased. If the respondents are asked, "Do you ever listen to *Spotlight Bands* sponsored by Coca-Cola on WJZ every night except Sunday, from 9:30 to 9:55 P.M.?" the percentage will jump again. This process can be pressed to the point where the respondent is given a guilty feeling if he fails to reply favorably. A fat figure, in short, can be developed by this manipulation of the roster questions. This factor would, of course, be expected to be a constant for all programs on the given roster, but it varies widely from one roster study to another, depending upon how each is conducted, and makes it difficult to compare the results obtained by one organization with the roster with those obtained by some other organization.

A third added factor causing variation in the roster method is the length of the list of programs used. By using a very long roster, relatively small ratings will be obtained. But the ratings will increase progressively in size as the length of the roster list is decreased. By reducing the roster to a few programs, magnificent ratings can be obtained. In fact, within limits, a rating of any desired size can be obtained merely by controlling the number

of program names printed on the roster and the amount of aid furnished in obtaining the recall.

C. THE UTILITY OF DAY-PART RECALL AND ROSTER METHODS

The above limitations apply to recall methods as a means of measuring program audience size. It does not follow that the methods are without any value whatever. Studies of recall methods made by the present writers reveal that, despite the inability to measure audience size accurately, they do measure with a fair degree of accuracy the relation of the size of the program's audience in one population sub-group to the size of the same program's audience in some other population sub-group. For example, it may be found in a roster study using a list of fifteen programs that twenty-five per cent of the families in large cities report listening to Eddie Cantor during a given broadcast and thirty per cent of the families in small cities and towns report listening to him. This does not mean that Eddie Cantor actually was heard by twenty-five per cent of the one group and by thirty per cent of the other. But it does mean that whatever the actual size of his audience was in the large cities, the size of the audience in the small cities and towns was twenty per cent greater or in the ratio of thirty to twenty-five.

This capacity of the recall methods to yield re-

liable relations between the audience size in one segment of the population and that in another is of inestimable value when these methods are used in combination with other methods which yield reliable measures of audience size in one segment of the population. Such combinations will be discussed at greater length in the final chapter.

D. IMMEDIATE RECALL

What has come to be known as the "immediate recall" method was developed by the Hooper organization. In this method, the respondent is asked what he is listening to now (coincidental) and what he was listening to fifteen minutes earlier (immediate recall).

When the method was designed it was believed that, because of the short interval of elapsed time between broadcast and recall, the method would free itself of most of the sources of variation which afflict other recall methods. It was found to be partially free of the influence of forgetting, but no recall method can free itself of the differences in the degree of impression made by different programs.

The method was therefore dropped by its originator as a device for measuring audience size. It does, however, retain a certain degree of usefulness as a very simple device for measuring the flow of audience to and from programs.

VIII. FIXED SAMPLE METHODS

Two general types of methods which employ fixed samples have made their appearance in radio recently: the mechanical recorder and the panel methods. While the former has received the greater share of attention, there is reason to believe that the latter will be developed and perfected in the near future and will come to occupy an important position in the study of radio audience characteristics. The mechanical recorder method is sometimes said to be a panel method. However, in marketing research, the field in which the technique originated, "panel members" are usually assumed to be human. The character of the panel method is indicated by the other names that have been applied to it—"Consumer Jury", "Case History", and "Diary".

The strengths and weaknesses of the mechanical recorder and of the human panel methods as ways of studying radio audience characteristics are so different that the methods should never be identified. One and only one group of problems they have in common: those arising from the fact that both use fixed samples. Each is therefore considered in a separate section in this chapter.

A. THE MECHANICAL RECORDER

The mechanical recorder method uses a "fixed sample" out of necessity. The recorder is an instrument attached to a radio set which makes a record on tape whenever the radio set is turned on. From the length of the lines drawn on the tape and their position the following may be determined: the time of day at which the set was turned on, the length of time it was in operation, and the stations to which it was tuned. The writers were among the first to undertake experimental investigation of the potentialities of the mechanical recorder method, and after two years of experimental work concluded that while the method had definite merit in some directions, it appeared to be unsatisfactory for the continuing, basic needs of the radio industry as a whole. At the present writing A. C. Nielsen Company, and Radio-Graph Corporation, among others are experimenting with the mechanical recorder.

The major problems characteristic of this method fall into two general categories: first, the strengths and weaknesses of the mechanical recorder itself as a way of measuring radio listening behavior, and secondly, the strengths and weaknesses of a "fixed sample" as a means of representing a population. The reported recorder experiments have concerned themselves largely with the former part of the problem. The second aspect of the problem, that created by the use of a "fixed sample", appears to have re-

ceived consideration only by individuals and organizations employing "panels" based on "fixed samples."

According to the "Nielsen Researcher" the A. C. Nielsen Company's mechanical recorder operation offers clients figures on the following:

1. Average hours of listening per day.
2. Average minutes of listening per hour.
3. Per cent of families using radio by hours of the day.
4. Program audience as a per cent of the total sample.
5. Program audience as a per cent of the sample using radios.
6. Average minutes listened per broadcast.
7. Rating expressed in terms of average audience.

Each of these seven is to be reported by the following audience breakdowns:

- a. Three income groups.
- b. Five sizes of locality including rural.
- c. Telephone and non-telephone homes.
- d. East and Central time zones.

Programs are analyzed in the following ways:

8. The internal pattern of listening.
9. Audience gain and loss in each minute.
10. Where the audience comes from (broken down by three categories).
11. Where the audience goes to (broken down by three categories).
12. Audience turnover.

13. Duplication of listening between programs where client has more than one program.
14. Frequency and length of listening.
15. Sales effectiveness.

In addition to the above information, the service offers station coverage data for individual stations. These consist of the following:

16. Per cent of families reached by station.
17. Per cent of total minutes spent listening to each station.
18. Average minutes of listening to each station by home listening.
19. Composition of the station audience by income groups, size of locality and telephone and non-telephone homes.

The sample used is described as "stratified" by the following: (1) number of radios in the home, (2) geographic areas, (3) size of locality, (4) income classes, (5) occupation, (6) race, (7) family size, and (8) telephone and non-telephone homes. Recorders are placed in 800 homes located in parts of the East and the North Central areas. Within this area are eight million radio homes—about one-quarter of the radio homes in the nation.

The mechanical recorder method carries a strong appeal to the imagination but, as in the case of any other method, its accuracy and utility to the industry can best be determined by examining it in the light of the six basic principles of radio audience research.

1. Does the Method Yield a Valid Measure of Radio Listening?

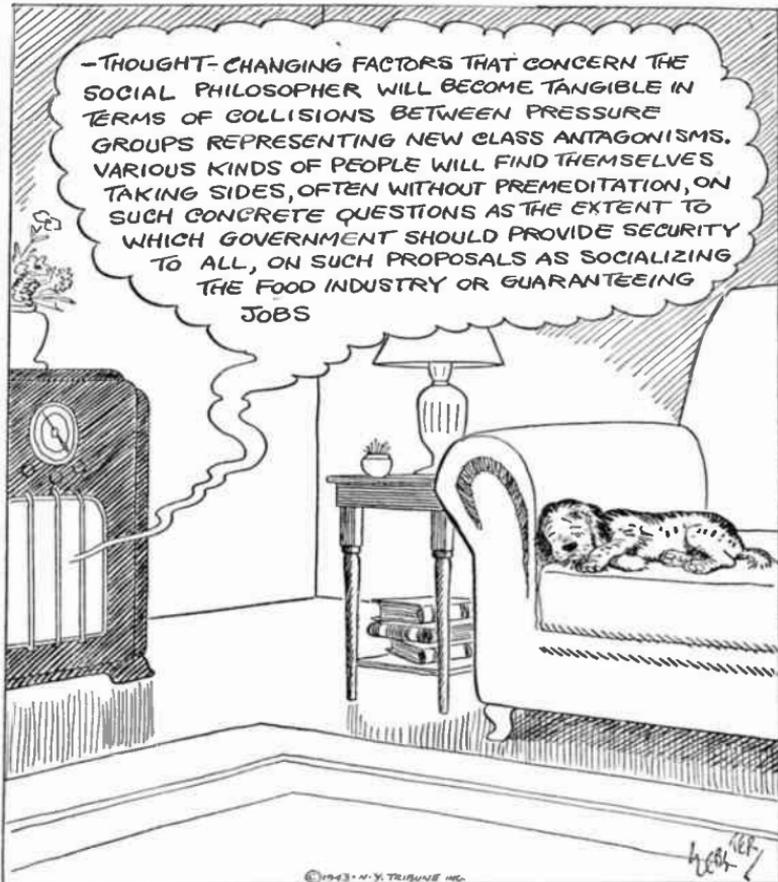
The recorder, when it is in working order,¹ yields a continuous record of the operation of the radio set. When the set is turned on that condition is recorded, together with the station to which the set is tuned. When the tuning of the set changes, the instrument records the change in tuning. That is to say, when it is in good working order, the recorder may be expected to yield a reliable measure of set operation and dial position. These are the only facts revealed by the method. The method reveals no factual data whatsoever on listening. A set may be tuned to a station for ten minutes or for twenty-four hours but such testimony as the recorder gives does not prove that someone in the family listened to the programs broadcast over that station for all or even more than a very small part of the time. These considerations are the more significant in the light of claims that the recorder method yields nothing but "facts". Concerning radio listening, it actually yields nothing but inferences. That there is some relation between set tuning and listening is in itself only an assumption, albeit a highly probable one.

¹ Neilsen reports that approximately ten per cent fail each month. Crossley's experience shows higher rate of failure. The writers' experience indicated also a large percentage of homes in which the investigator was unable to retrieve the tape because occupants were persistently "not at home."

It seems obvious that some relation between tuning and listening must exist but the important ques-

THE UNSEEN AUDIENCE

OCT. 27



tion that has to be answered for each program is: "What is the relation for this particular program?" To assume on the basis of the evidence at hand that

the relation is one of perfect correlation, is sheer absurdity.

Archibald M. Crossley of the Radio-Graph Corporation states in *Advertising and Selling*, July, 1939, that the relation is far from perfect. As a result of one of many recorder studies in which he compared tuning with listening, Crossley reports: "We found as high as 20-25 per cent of the sets in operation for periods exceeding ten minutes when no one was in the room with the set."

Anyone who takes the trouble to note for a few hours in his own home the relation between listening and set operation will confirm Crossley's results. A real difference between tuning and listening will be found even in the absence of any prolonged periods of tuning to a single station. Add to these considerations the facts revealed in Charts XXXV and XXXVI and it becomes clear that the relation between tuning and listening is far from one of perfect correlation.

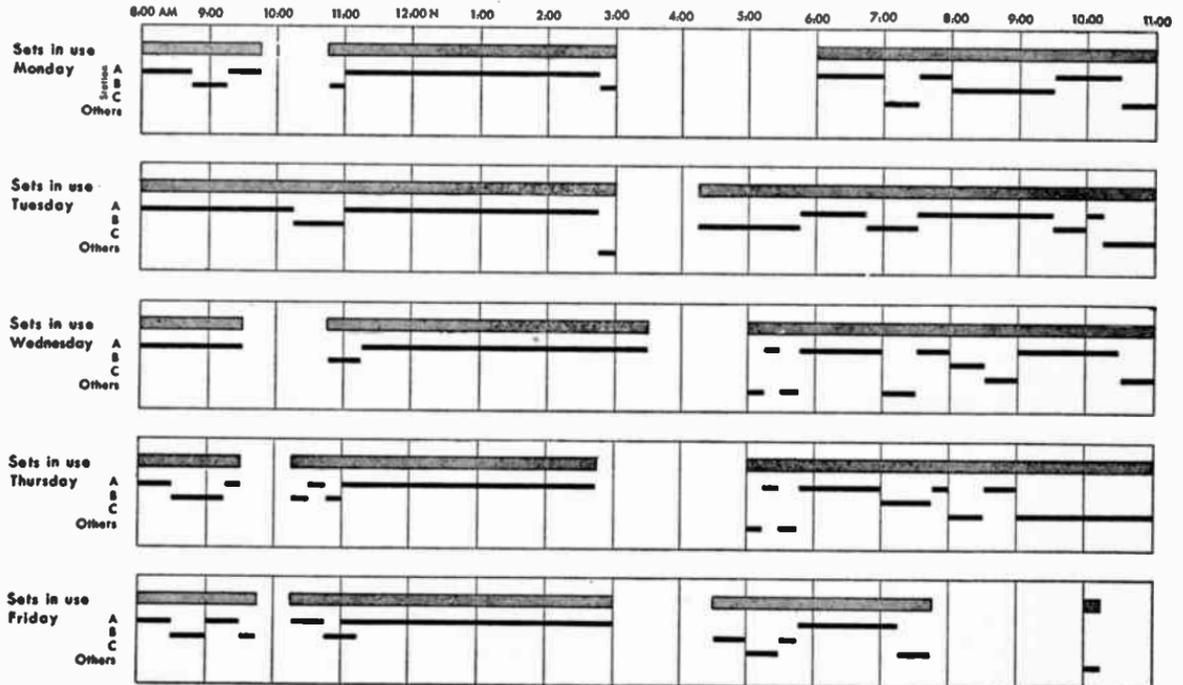
Chart XXXV, from a study by Crossley, presents the testimony furnished by one set for one week, Monday through Friday, 8:00 A.M. to 11:00 P.M. Analysis of this record reveals the following: In only twenty-eight per cent of the tunings did the set remain tuned to one station for more than an hour. However, when "Tuning Time" rather than "Number of Tunings" is used as the basis of comparison, it is found that this set was tuned to one station for an hour or more during sixty-three per

CHART XXXV

EXAMPLE OF A SET USER'S STATION TUNING

MONDAY-FRIDAY FROM 8:00 AM-11:00 PM AS FOUND BY RECORDER

Source: Crossley, Advertising and Selling, October 1940



cent of the time it was in operation. And during thirty-two per cent of the time it was in operation, it was tuned to one station for three hours or more.

Chart XXXVI is reproduced from another of Crossley's¹ studies on the recorder. It illustrates types of set-tuners revealed by the recorder tapes. As Crossley indicates, it is highly improbable that there is a close relation between tuning and listening in the "Let-It-Ride" type. In the case illustrated, the set was tuned to one station continuously throughout a six-hour period.

Periods of five to six hours of tuning to one station may occur in only a small percentage of the cases, but the importance to recorder measurements of even a single instance is more readily apparent when it is realized that one five-hour "tuner" as recorded by the mechanical recorder is equivalent in listening time to 300 one-minute tuners, 150 two-minute tuners, 100 three-minute tuners or 60 five-minute tuners. And when the sample is only 450-700 homes, the influence of even one "Let-It-Ride" tuner may cause appreciable distortion.

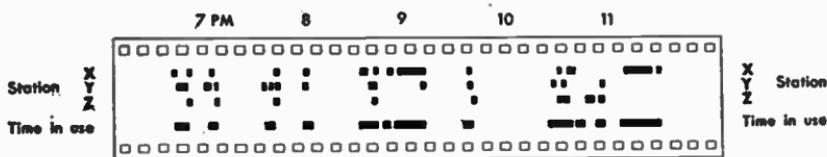
It must be concluded from these findings that some relation between set tuning as measured by the recorder and listening behavior undoubtedly exists. But it is also apparent that the degree of relationship and its variation from one program to another or one time of day to another remains quite unknown. Whatever else the relationship may be,

¹ *Advertising and Selling*, October, 1940, page 24.

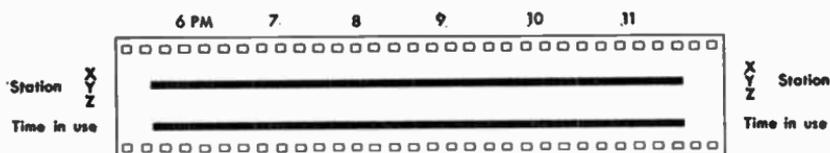
CHART XXXVI

SAMPLE RECORDER TAPES SHOW TYPES OF SET-USERS

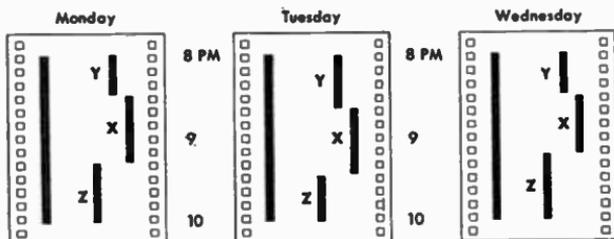
Source: Crossley, Advertising and Selling, October 1940
Tapes from Radio-Graph Corporation



"Restless Type"- Shifts stations often, probably attentive, but misses much of programs, commercials

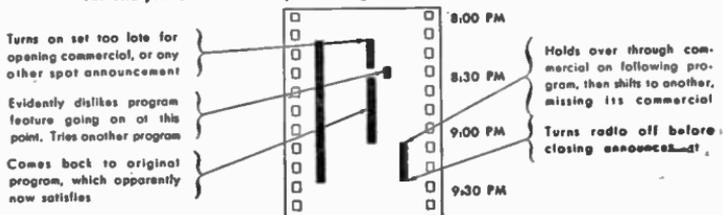


"Let-It-Ride Type"- Runs sets for long periods on same stations; degree of attentiveness probably low



"The Middle Type"- Tunes sets much the same way each day; attentiveness undoubtedly varies widely

An analysis of recorder tape showing set in use from 8:04 to 9:27 PM



clearly it is not one of perfect correlation. Until this relationship becomes known for each individual program the mechanical recorder method cannot yield a valid measure of radio listening.

2. Does the Method Yield a Valid Measure of Program Preferences?

The determination of program preferences is a fundamental purpose of audience size measurement, but it cannot be assumed *a priori* that any technique designed to attain this end does so automatically. In the recorder technique no testimony is obtained from the listener. The only data from which program preferences may be judged is a line on a tape. How long must the line be before it indicates a preference for a program on the part of the set-tuner?

Reconsider the case of the man who, searching for news, punches six buttons on his radio, one after another, listens momentarily to the sounds coming from each station and then turns off the set. Is he expressing a preference for the programs on each of the six stations to which he dialed? Obviously not, despite the fact that the recorder might register tuning to six stations.¹

The primary function of sampling research in radio is to determine program preferences by count-

¹ This is not true of the recorder used by Radio-Graph Corporation which records no station tuning of less than twenty seconds in length.

ing the number of homes which "vote for" each program by listening to it through choice. In the above illustration, the man did not listen by choice and thereby "vote for" each of the six programs he heard. He chose none of them. Rather, he "voted against" them by dialing them out as soon as he perceived their content.

But if a few seconds on the tuning record does not indicate a "vote for" a program, does a minute? Two minutes? Three minutes? Five minutes?

If one tunes in on a drama which is already in progress, listens for three minutes to pick up the theme, and then tunes it out, is he "a listener"? Does the three-minute line on the tape reveal that he has a "preference for" the program? If he picks up the story in one minute and then tunes it out, is his "preference for" the program any less than would be the case if it had taken him three minutes to perceive the program situation? The recorder takes no testimony from the tuner who is constantly casting votes both for and against programs. It is therefore necessary to draw inferences regarding the meaning of lines found on the tape—to choose arbitrarily some length of line and declare that everybody whose radio was tuned to a given station long enough to make a line of that length expressed a preference for the program.

Obviously there is no point in time and no length of line that is any more justifiable for arbitrary selection than any other. Any length of time se-

lected to indicate "a listener" involves an appreciable error. If short periods are chosen, many respondents are included as "listeners" who actually "voted against" the program. If a long period is selected, many respondents will be excluded who listened by choice for as long as they were able. An intermediate time will result in both kinds of errors.

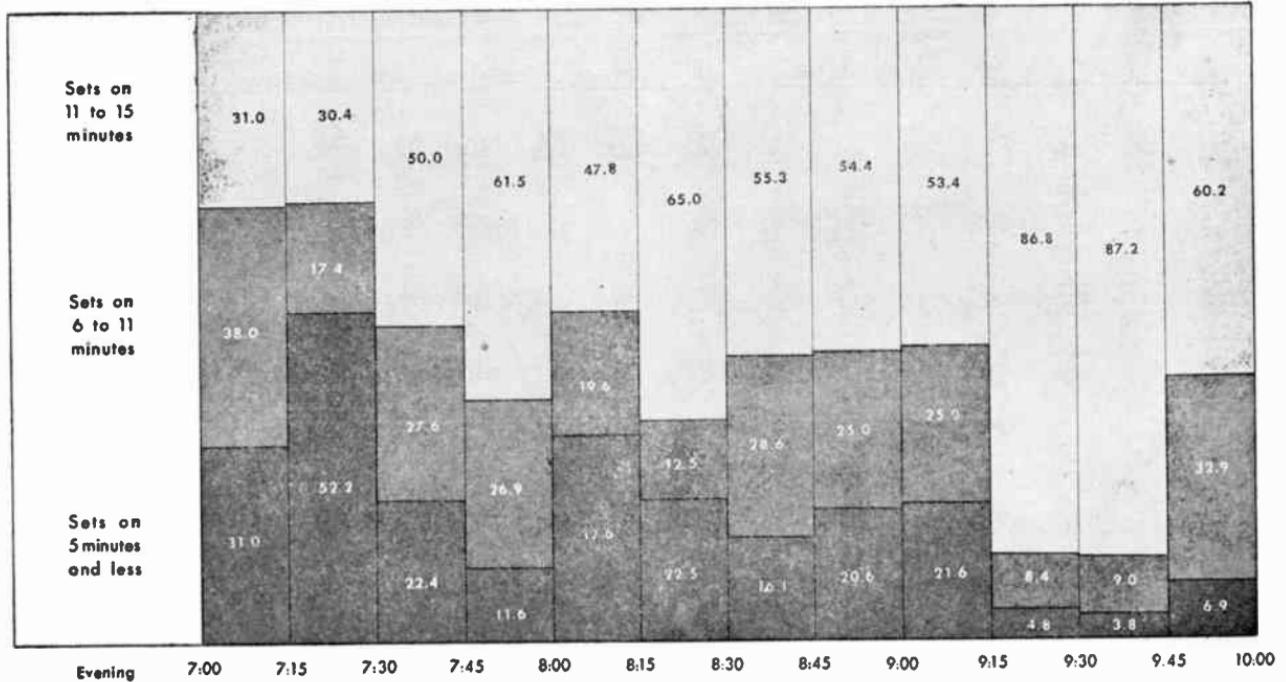
Furthermore, any length of time-line arbitrarily selected will not apply equally to all programs; for it takes longer to perceive the content of some programs and to come to a decision to "vote for" or "vote against" them than it does for others.

Evidence which supports these considerations is found in Chart XXXVII reproduced for Crossley's 1940 article cited above. This chart shows that in the early periods of the evening, tunings of five minutes or less in length represent 52.2 per cent of all tunings. What part of this 52.2 per cent represents "votes for" and what part "votes against" the programs covered? One guess is as good as another. This chart also demonstrates that the per cent of tunings of five minutes or less in length varies widely from one time period to another. While the high is 52.2 per cent from 7:15 to 7:30 P.M. the low is 3.8 per cent for 9:30 to 9:45 P.M., a range of variation from one time period to another of 48.4 per cent of all tunings.

In view of the inability of the recorder to determine which lines on the tape indicate "votes for" and which "votes against" a program, it must be

CHART XXXVII
DISTRIBUTION OF EVENING TUNING INTERVALS

Source: Crossley, Advertising and Selling, October, 1940



concluded that the mechanical recorder method does not yield valid measurements of program preferences.

3. Is the Unit of Measurement Standard Under All Conditions?

The recorder method reports program ratings primarily in terms of average audience. In following this practice, the attempt is made to maintain a unit of measurement which is constant under all conditions.

To get a measure of average tuning all of the time lines indicating tuning to a given program would be added together and divided by the number of cases. But Crossley has shown that short time tuning, much of which may represent "votes against" a program (how much is unknown), varies within wide limits. He has also shown that the relation between listening and tuning is far from perfect. Since the real significance to listening in the case of each line is both unknown and variable, the average obtained from them is also an unknown and a variable.

The Greek philosopher Parmenides remarked that an error is not removed by multiplication.

It seems necessary to conclude from these findings that the unit of listening measurement employed in the mechanical recorder method is not standard under all conditions.

4. Is the Method Equally Applicable to All Broadcasts and All Populations?

It is in the universality of application that the mechanical recorder attains outstanding value. As has been pointed out above, the coincidental method is not applicable to programs broadcast before 8:00 A.M. or after 10:30 P.M., local time. The day-part recall method, while it is applied to programs broadcast after 9:00 P.M., studies these programs the following morning, thereby introducing a longer recall interval than is employed for programs broadcast between 9:00 A.M. and 9:00 P.M. The mechanical recorder, on the other hand, runs day and night and records tunings and set operation at all times.

Similarly, the mechanical recorder method is equally applicable to all populations who have radios in their homes. This characteristic of the method permits analysis of data by race, occupation, economic status, educational level and telephone ownership which can be obtained by neither of the telephone methods considered above. It also furnishes the basis for combinations with other methods which will be considered in Chapter IX.

5. Is the Sample Representative of the Population from Which It Is Drawn?

In considering the degree of representativeness attained in a sample, the purposes for which the

study is designed must be considered. One of the declared purposes of the Neilsen radio audience measurement service is to report findings on small segments of the total sample such as the number of homes listening to each radio station. This requires that a sample be developed which is internally consistent throughout all its parts. The second purpose is to measure the total number of homes listening to a program. A third is to measure the entertainment or attraction value of network programs. That is to say, to furnish network program ratings.

The representativeness of the sample would have to be judged by considering each of these factors individually.

Since the writers know little of the actual distribution of current recorder samples, little can be said on this score.

Internal consistency between all of the parts of the sample is reported by Neilsen to be obtained through "scientifically controlling" it in eight dimensions: (1) number of radios in the homes, (2) geographic area, (3) size of locality, (4) family size, (5) race, (6) occupation, (7) income status, and (8) telephone ownership.

In sampling parlance a "controlled" sample is usually interpreted to mean a "stratified" sample. But, as has been pointed out by Samuel E. Gill the Neilsen sample cannot be "stratified" in eight dimensions because it is mathematically impossible to make so many stratifications with a sample of 800 homes.

It was shown in Chapter III that in "stratifying" a sample the number of population sub-groups within which random sampling must be conducted increases as the multiple of the subdivisions of each condition stratified.

Table XIV shows the calculated population subdivisions that result from stratification in the eight dimensions said to be "scientifically controlled" in the Neilsen sample. As Gill has pointed out in his memorandum, the largest of the subdivisions would contain some hundreds of times more homes than the smallest. If the smallest subdivision was represented in the sample by one home the largest subdivision would require many hundreds of homes. The sample stratified in these eight dimensions would necessarily attain a size of some hundreds of thousands of homes.

TABLE XIV

Population Subdivisions Resulting from Stratification in Eight Dimensions (From a Memorandum by Samuel E. Gill)

Breakdown	Number of Divisions	Cumulative Total Subdivisions
Income Class	8	8
City Size	5	15
Geographic Area	2	80
Family Size	5	150
Telephone and Non-Telephone	2	300
Occupation	5	1500
Number of Radios in Home	3	4500
Race	4	18000

For purposes of yielding the internal consistency essential to reporting on small segments of a sample "stratified" in eight dimensions, 800 homes—or 8,000 homes—are patently absurd.

a. *The need for knowledge on the number of stations habitually heard.* As was pointed out in Chapter III, one fundamental dimension of stratification which is necessary in making accurate measurements of "total homes listening" is the proportion of the families in the population which habitually listen to one, two, three, four, five, or more radio stations. Without accurate representation of homes in each of these subdivisions, no small sample designed to measure listening in the total population can represent listening conditions validly.

For example, a daytime program may get a rating of 1.0 in areas served by many stations but in areas in which only one or two stations may be listened to in the daytime, the program may get a rating of from 10.0 to 16.0. This example is not based on supposition but on conditions repeatedly encountered in intensive studies of listening areas. The unvarnished truth of the matter is that no organization in the country knows how to lay out a sample that represents the number of stations listened to habitually in the total population. There are only two sources of data from which this could be determined. They are the all-country coverage studies made by two major networks. And they do not agree with each other.

b. Representativeness and the fixed sample. Quite apart from the above considerations there is the whole area of unknowns concerning representativeness of any fixed sample. First, there arises the question of bias. Do families which permit recorders to be attached to their radios represent a selection which introduces a bias? It definitely introduces bias in some directions. For example, it is highly improbable that the mechanical recorder could be placed in any home which habitually listens to German shortwave broadcasts. More important still is the question of whether or not the recorder introduces a bias in the selection of broadcast band stations and programs. While no data are currently available on the situation, it would be not at all surprising if "prestige" programs such as concert and symphonic music took quite a jump in homes in which the recorders were located.

Secondly, there is the question of how long a "representative" fixed sample represents what it is supposed to. Obviously, it can be representative of the population from which it is selected for only a very short time. To the extent to which it remains "fixed", it loses its representativeness rapidly because both the family and the population change. New family members are born, others die, and all grow older day by day. With change in age comes changes in educational level and standard of living. Each family that furnishes data for any significant period of time can, of necessity, not represent con-

ditions it was originally selected to represent. Cawl¹ has remarked in this connection that there are as yet "no standards by which to judge when the internal changes in the composition make it (the fixed example) obsolete".

6. Is the Sample of Sufficient Size to Yield Figures That Are Statistically Significant?

It was remarked in Chapter IV that the size of the statistical error in a sampling operation depends alone upon the number of cases studied. The size of the statistical error has absolutely no relation to the question of whether the sample is so distributed as to represent the population from which it is selected. That is to say, the method might be entirely incapable of yielding a valid measurement of actual listening behavior; but this fact would have nothing whatsoever to do with the size of the statistical error involved. The latter is the error that is inherent in the sample size over and above the errors of representativeness of the sample and the limitations of validity of the method.

For example, the statistical error of figures based on the total sample of 700 homes would be satisfactorily small for many purposes. However, if this sample is broken down into very small segments,

¹Some of the problems and needs for experimentation through which standards can be developed which are created by the use of fixed samples have been discussed with rare insight by Franklin W. Cawl in the *Journal of Marketing*, July, 1948.

the findings obtained on each might have very low reliability. In order to determine the statistical adequacy for each breakdown, each part of the operation must be considered individually. In these considerations it will be assumed that the recorder sample is 700 homes.

a. *Measurements of "Total Homes Tuning"*. "Total Homes Tuning" figures would be based on the total sample of 700 homes. Chart III, page 57, shows the statistical tolerance for a figure of 10.0 based on this sample would be ± 2.2 . That is to say, any comparable figure which was less than 7.8 or more than 12.2 would be significantly different from the obtained 10.0. This sample would be considered satisfactory for many purposes in radio.

If this sample of 700 homes is broken down by telephone and non-telephone homes, three economic levels, and five sizes of locality, the reliability of the figures decreases.

Assuming that forty per cent of the families have telephones, the sample representing telephone homes becomes about 280 families. A figure of 10.0 based on a sample of this size would have a maximum expected variation of about ± 3.4 .

If the income categories employed are of equal size, the sample representing each would be about 235 homes. The significant difference for a figure of 10.0 based on this sample would be ± 3.9 . That is to say, if "News of the World" got a rating of 10.0 in the middle income group, its rating in the lower in-

come group would have to be 6.1 or less to be significantly lower and 13.9 or more to be significantly higher.

If each of the five categories of city size employed were equal (which is impossible), each category would have about 140 homes. The maximum expected variation for a figure of 10.0 based on this sample would be ± 5.2 .

b. Program analysis. When analysis of the conditions internal to the program are considered, the recorder sample is 700 homes. Assume that ten per cent of the sample tuned in on a given program some time during its broadcast. The basic sample is 700; so the total number who "tuned in some time" is seventy. This is the sample size on which must be determined such characteristics as the audience flow.

If, in determining where the audience comes from, this sample were broken down into three subdivisions of source: (a) from some preceding program (on any network or local station), (b) from competing programs, and (c) from "off", and the three subgroups are of equal size, the sample for each would be twenty-three homes. These twenty-three homes would be used to determine from which particular network the listeners flowed "to the program". The same would hold for determining to which networks the audience flowed "from the program".

c. Sales effectiveness. In studying the sales effectiveness of a program based on the total sample of

700 homes, the following conditions obtain. Assume that ten per cent of the sample is found to have tuned to a program. The sample of listeners upon which the sales effectiveness study would be based is then ten per cent of 700, or seventy homes. If the client has a popular brand of a popular product, his brand may be found in thirty per cent of the tuning homes, or twenty-one homes. If the brand advertised competes with many other brands of the same product, as in the vitamin field, the advertiser's brand may be used in not more than eight per cent of the listening homes. In this case, results would be based on a sample of six user homes. Under these conditions, a brand shift in one home can cause an apparent change in the sales effectiveness of the program from five per cent to eighteen per cent. These are also the diminutive samples from which the client must judge the relative use of his brand among listeners and non-listeners.

d. Station coverage. There are in the coverage area claimed by WAVE in Louisville approximately 128,000 radio homes. If 2,600 recorder homes were used to measure listening nationally, eleven recorders would be located in the WAVE area. This number would be divided up by mileage zones, by income groups, city size, and telephone and non-telephone homes. Cities of each size would, at best, be represented by about three recorders and each economic level by four. And these, presumably, would yield a measure of the multiplicity of variables which

constitute coverage. Obviously a sample size of such order would lead to statistical absurdity.

e. Other limitations of the recorder method. Quite apart from the error involved in determining which tuners are and which are not listeners, and quite apart from limitations imposed by sample size, the recorder method of obtaining data is subject to certain other important limitations.

A fundamental limitation of the method is its inability to determine the size of the available audience during the broadcast of a program. The available audience for any program, as was pointed out in an earlier section, is the per cent of the sample of homes in which someone is at home and awake during its broadcast. If the family is asleep, at the movies, or at work in war factories, it is not a part of the available audience of any program broadcast during that time.

There is no way to determine from the data collected with the recorder whether the family was at home or away from home, awake or asleep. When the set is turned off, the recorder tells nothing.

Without data on the size of the available audience it is impossible to analyze satisfactorily the factors that may influence changes in the audience to a program. One or two examples will serve to indicate the necessity for such data. In April, 1942, (see Charts XII to XIV) most programs showed a greater drop than the expected seasonal declines. Why? Without a measure of the change in size of the

available audience it could not have been determined that putting the nation's clocks on War Time in February had advanced the seasonal decline in ratings by about a month.

A program broadcast on Sunday night may get a rating of 25.0. Another program broadcast on Monday night may get a rating of 27.0. Which program has the greater attraction value? All that is revealed by these figures is that the Monday night program attracted more people in the total sample. But in order to determine which program recruited listeners more efficiently, it is necessary to know the size of the available audience at the time of each broadcast. On Sunday night the available audience is at a minimum; whereas, on Monday night it may be expected to be maximum for the week. The Sunday night program which obtains a rating of 25.0 based on the total audience may recruit thirty-five per cent of the available audience while the program which rates 27.0 on Monday may recruit only thirty-three per cent of the available audience.

The difference in available audience from day to day, season to season, hour to hour, economic level to economic level, and city size to city size are so great that any attempt to compare program appeal from one population sub-group to another may lead to grave errors in judgment where no evidence concerning the available audience is at hand.

The subdivisions of the population most important to radio broadcasting are those of sex and age

of listeners. This follows from the fact that the majority of programs are designed to appeal to a specific sex or a general age level. Nearly all daytime serial dramas are directed to housewives. Some of the late afternoon and early evening shows are designed primarily for children. Later evening programs are, for the most part, designed for adults of both sexes; but a few sponsored by cigars and pipe-tobaccos are interested primarily in the male adult audience.

Because the recorder can obtain no testimony from the listeners, it is quite incapable of developing data which reveal the number of listeners per set during the broadcast of a client's program or the age and sex of the listeners.

As was pointed out in the discussion of the coincidental method, surprising facts concerning the appeal of programs are revealed by these analyses which can only be made accurately through the use of the coincidental method. There the audience is counted, family member by family member, while the program is being broadcast.

7. Strength of the Mechanical Recorder

While it is apparent from the above considerations that the mechanical recorder probably does not yield accurate audience size figures, it does not follow that the method is without great value. It seems highly probable that the method does show relative dif-

ferences in audience size in the various sub-groups of the population. This being the case, the mechanical recorder method may be used in combination with the coincidental method to obtain accurate size measurements for each of the population sub-groups. The operation of this combined method will be discussed in the final chapter.

The mechanical recorder is capable of gathering some data which cannot be obtained by any other method and it yields more reliable measures of some audience characteristics than can be gained by other methods.

First, it offers something of a measure of public dissatisfaction with radio offerings. This dissatisfaction would seem to be indicated in the short-time tuning recorded on the tape. Some experimental work would require to be done to determine the optimum length of line which indicates a "vote against" a program and how it varies for programs of different length and type, but this is no super-human job. And it would furnish both networks and advertisers with a completely new and most valuable tool—a measure of dissatisfaction with programs. The mechanical recorder also provides the best method for obtaining the flow of tuning from one program to another. The fact that the current recorder samples may be much too small to provide such data does not mitigate against the mechanical recorder as an ultimate method of obtaining them.

Similarly, the method provides excellent data on

the duplication of tuning between programs where the client has more than one program on the air. The current samples may be too small to yield significant data on duplication but the mechanical recorder is capable of yielding it, given an adequate sample. The same is true of frequency and length of tuning. And finally, if a sample of mechanical recorder homes were ever developed which was sufficiently large to measure "total homes tuning" locally and throughout the nation—that is to say, a sample which was laid out in accordance with radio's primary variable, station competition—the industry would automatically be furnished with the best possible measurement of station and network coverage. As was indicated in an earlier chapter, this sample would necessarily be very large, probably not less than 100,000 homes.

B. LISTENER PANELS

The panel method has come to the radio audience measurement field from that of marketing research. Even in the marketing field the method is a newcomer, the oldest panel operation being that conducted under the direction of Ray Robinson for *Woman's Home Companion*.¹ Anyone who works with a panel must become enthusiastic over its potentialities, but everyone is also left with the question in his mind of just what his results mean.

¹ For a discussion of this panel operation, see the article by Ray Robinson in *Advertising and Selling*, June, 1948.

Perhaps the best evaluation of the continuing panel technique is that presented by Franklin R. Cawl in the article cited above.² In this article, which should be studied by anyone interested in "fixed" sample operations, Cawl outlined the needs for study to develop standards of panel operation. While there is as yet no broad radio audience measuring service which employs the panel method, it seems probable to the writers that when the standards are finally developed, the method will come to occupy an important position in the field, yielding much needed data.

One of the first radio advertisers to employ the panel method in the study of commercial content and structure was Du Pont. In these studies, the form, appeals and memory value of "Cavalcade of America's" commercials were studied intensively with the result that a commercial formula was designed which held the audience and which was remembered by the listeners on the day following the broadcast as well as was the program content. Stanton, Churchill, and Smith of Columbia Broadcasting System and Samuel Barton of Industrial Surveys Co. have used the method most effectively in the study of program type, form, and theme.

The essence of the panel method resides in the fact that it furnishes a relatively "fixed" sample of homes in a population which can be employed for

² *The Journal of Marketing*, July, 1948.

continuous study over a long period. Insofar as the composition of the panel remains relatively constant, it permits the determination of the growth and decline of a wide range of attitudes, opinions, preferences and behavior and their study in the light of the factors influencing the population such as radio programs heard, newspapers and magazines read, educational level, family size, etc.

In most forms of the panel method, the members keep a record or "diary" of certain of their activities. For example, in a panel designed for the study of radio programs and their effectiveness, the panel member might keep a record of (1) all radio listening by time, station and program name; (2) all time periods when no one in the family was at home; (3) the time at which each family member went to bed at night and got up in the morning; (4) all products bought (recorded by brand name) and moneys expended. From these data it would be possible to obtain measurements of the following:

1. Program audience size (relative).
2. Duplication of listening between programs.
3. Frequency of listening to any program (program loyalty).
4. Audience flow.
5. Relation of program listening to use of sponsor's product.
6. Product satisfaction and product "flow".
7. Growth or decline of interest in program.
8. Hours of listening per day by family size.

9. Available audience at any time (relative).
10. Composition of program audience in terms of number, age, and sex.
11. Relative program appeal in population sub-groups such as educational and economic level, localities of different size, etc.

In addition to these, many other records could be obtained through interviewing the panel members. For example, the developments of attitudes significant to concurrent national conditions and the relation of these to specific program listening could be followed. Similar studies might be made of attitudes toward large companies and of the effectiveness of each company's radio efforts to build good will. Supplementary interviewing of the panel members would also offer an excellent method of studying the form, placements, content and memory value of commercials and specific elements of program content.

Even in its early stage of development some of the limitations of the method as it applies to radio audience characteristics reveal themselves. First, it becomes clear that the panel method resolves into a "memory method". Under ideal circumstances listeners would record their listening, the minutes away from home, the time of going to bed and getting up, brand purchases and money expenditures as they occur. However, in panel operation as in the science of physics, "ideal conditions" are purely imaginary. They do not exist in nature. This holds

with particular force where human nature is concerned.

Because of the conditions of living, of the need for making records in panel operation and of the human tendency to resist interruption of activities in progress at the moment, the panel resolves itself into a memory method in which the memory influences are indeterminable. Sometimes the panel members may record events as they happen but because Americans are very busy people, the records are prone to be made at some interval of time after the event to be recorded has occurred. The interval may be a few minutes or it may be many hours. There is no way of knowing in a panel study. The interval of elapsed time is an indeterminable element but one which involves memory.

As a result, panel measurements on radio audience size may be expected to resemble those obtained by other recall methods in that they would not reveal actual audience size for any broadcast. They would, however, be expected to reveal *relations* between the size of the listening audience in the different population sub-groups of the sample. As in the case of the other recall methods and the mechanical recorder, this limitation may be overcome by using the panel in combination with the coincidental method. The combination would be expected to yield reliable audience size measurements for a cross-section of a population or for any of the population sub-groups.

Secondly, the problems of representing radio listening homes in the nation that apply to the mechanical recorder method also apply to the panel method. In addition to the problems created by the fixed sample there remains the stark fact that at present the standards of panel operation, the limits of its utility and its accuracy must still be determined. Despite its limitations, however, it seems probable to the writers that this method has bright prospects for future development. The manner in which it can be used in combination with the coincidental method to yield a wide range of highly reliable data will be discussed in Chapter IX.

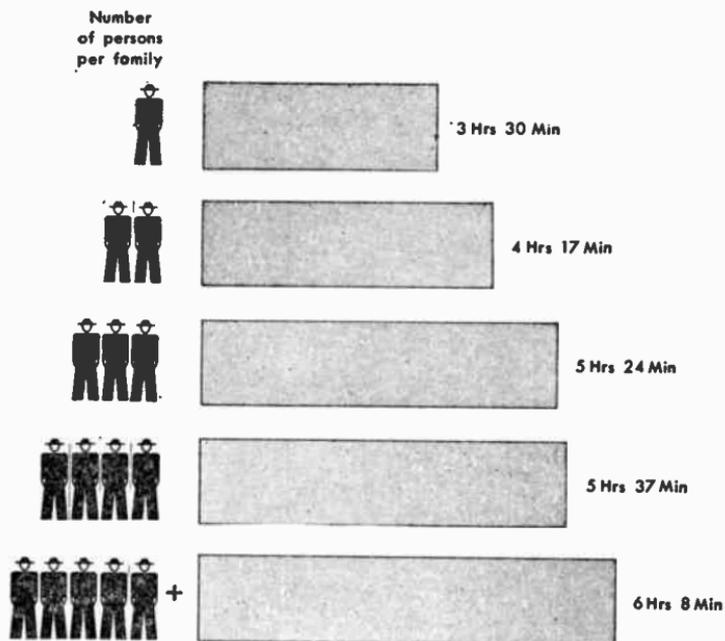
The "Diary" method. A variation of the panel method recently developed out of research conducted by Churchill and Smith of Columbia Broadcasting System is that known as the "Diary" method. In this method the panel is drawn by mail from a preselected sample and reports are made by mail. The prospective respondent is furnished with seven sheets designed for mailing upon which listening and station call-letters are recorded by quarter hours. The studies were made, for the most part, within a given station's area.

The method is designed for single studies of one week in length rather than for continuous operation. The prospective respondent is given inducement to co-operate in the form of a gift. There is an advantage here over many panel operations in that this method does not use the panel long enough

for it to lose its representativeness. It has the disadvantage that it furnishes no long term trends—a fundamental advantage of some other forms of the

CHART XXXVIII
HOURS OF LISTENING BY FAMILY SIZE

Source: Columbia Broadcasting System Diary Study - 1943



panel method. The "diary" method was designed primarily to develop data not furnished by any of the national audience services. It has, in fact, yielded for the first time some data which have long

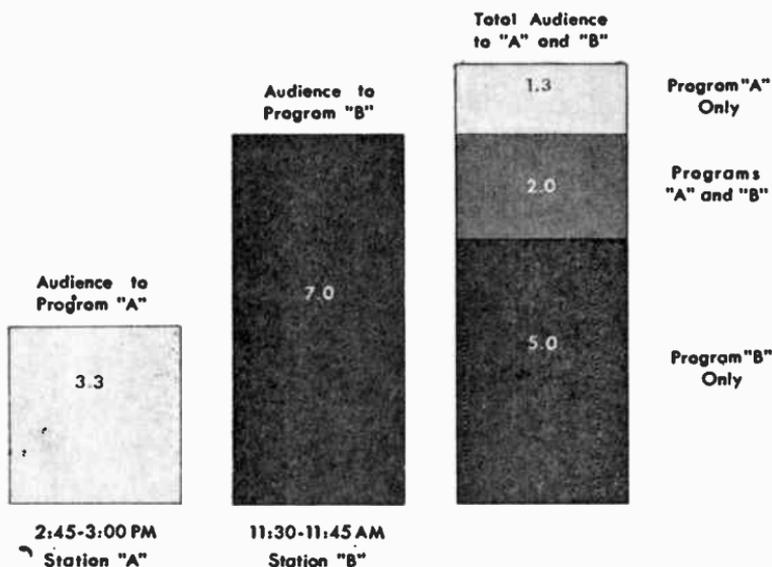
been needed by the industry and by samplers. The minimum sample size used in each local study was 750 homes.

CHART XXXIX

DUPLICATION OF AUDIENCE

TWO DIFFERENT PROGRAMS BROADCAST BY THE SAME SPONSOR ON THE SAME DAYS OF THE WEEK

Source: Columbia Broadcasting System Diary Study- 1943



An outstanding example of contributions that may be made by the "Diary" technique is illustrated in Chart XXXVIII which shows the relation between hours of listening per day and family size.

The trend leaves little doubt that listening time is related directly to family size. This chart probably enjoys the distinction of representing the first actual measurement of this relation.

Chart XXXIX, also taken from the report of the "Diary" study, shows how the method may be used to measure duplication of listeners between two programs presented by the same advertiser. It indicates that by adding Program A to Program B the advertiser reaches only nineteen per cent more homes than he reaches with Program B alone.

Quite as important as their contribution to knowledge of radio audiences is the fact that the CBS studies have contributed greatly to the determination of some of the standards of panel operation and application.

IX. COMBINATION OF METHODS

Data which reveal with uniform accuracy and reliability listening conditions (a) on a national scale, (b) within a station listening area, and (c) within individual localities have been and remain a basic need of the radio industry. It has been shown in the preceding chapters that none of the methods currently in use in the field of radio audience measurement is capable, in and by itself, of satisfying all of these needs. Each of the methods has been examined and found to be subject to important limitations. But the matter does not end there. One possibility remains to be considered: Can a combination of methods be developed in which one method will compensate for the other's limitations and thereby satisfy the overall needs of the industry?

The use of combinations of methods in the sampling field is not new. Gallup's famous prediction of the error in the *Literary Digest* poll in 1936 was based on a combination of personal interview and mail methods. Blankenship,¹ research director

¹ Blankenship, Albert, *Consumer and Opinion Research*, 1948, p. 58.

of N. W. Ayer and Son, has concerned himself with similar combinations of samples in the marketing field. The product use study designed by C. E. Hooper, Inc., which was described in Chapter V, involves a combination of methods.

Archibald Crossley,² who has had wide experience with radio audience measurement methods, recognized the need for a combination of methods in radio audience research. He has expressed the view that results most favorable to the industry will accrue from the use of the day-part recall and the mechanical recorder methods in combination.

This combination has much to recommend it but it is subject to such marked limitations that it could not possibly furnish all the basic data required for a thoroughly rounded service.

In the first place, the whole purpose of using methods in combination is to compensate for limitations inherent in each. In the case of the day-part recall and the mechanical recorder methods, both share important limitations in common. The most important of these is the basic consideration that neither yields an exact measure of either listening behavior or of program preferences. As was shown in Chapter VII, the day-part recall method is prevented from measuring listening accurately because it is subject to variable memory influences and to a constantly shifting base.

It was shown in Chapter VIII that the mechani-

² *Advertising and Selling*, Oct., 1940.

cal recorder method is similarly limited because many of the short lines on the tape actually indicate "votes against" a program rather than program listening, while in the case of the longer lines there is no way of knowing what part of the time someone was listening. There is, therefore, no compensation for the primary limitation of each method in this combination.

Secondly, neither method is capable of measuring the available audience or of yielding measurements based on it. The day-part recall method employs a sample of homes in which someone is at home at the time of the interview. It is quite incapable of determining what part of the population was at home during the broadcast of any given program. The mechanical recorder method, on the other hand, employs a fixed sample which purports to represent "total radio homes" in the population. The only record obtained is one of set operation. When the set is turned off, the family may be awake or asleep, at home or away on a vacation. There is no way of measuring the size of the available audience by the mechanical recorder. Thus the two methods fail to compensate on this score. Thirdly, neither method is capable of yielding data on the composition of the listening audience in terms of number, age and sex of listeners. Instead of compensating for each other's limitations, therefore, a combination of the day-part recall and mechanical recorder methods duplicates the fundamental limitations of each.

The preceding chapters have shown that the coincidental is the only one of the methods considered that yields an accurate measure of either audience size or program preference. Since these measurements are fundamental requirements, it follows of necessity that the coincidental method must furnish the base of any combination of methods which will satisfy the needs of the radio industry. The question of the selection of combinations of methods therefore resolves itself into the following: Which other method may be used in combination with the coincidental to yield the greatest amount and the most reliable of the data required on radio audience characteristics.

Consider first the limitations of the coincidental method for which compensation is required.

1. For economical operation it requires a telephone sample.
2. Telephone results cannot be analyzed for educational, occupational or economic status.
3. The method reveals nothing concerning the flow of audience to and from programs.
4. The method reveals nothing concerning the duplication of listening where the client has two or more programs.
5. The method is applicable to programs broadcast between 8:00 A.M. and 10:30 P.M. only.

B. COMBINATION OF THE COINCIDENTAL AND PRINTED ROSTER METHODS

It was shown in Chapter VII that the printed roster method does not yield measurements of actual audience size. It was also pointed out that the

method does yield one highly important piece of information. It measures the *relative* audience size in the different sub-groups of a population. For example, if fifteen per cent of the homes in large cities and thirty per cent of the homes in small cities report having listened to a program in a roster study, it does not mean that fifteen per cent of the one group and thirty per cent of the other actually listened. But it does mean that, whatever the actual size of the listening audience in large cities may have been, the listening audience in small cities was twice as great.

The roster method yields a reliable relation between the size of listening audience in sub-groups of the population. This is a fact of first importance, for it furnished the basis for compensation for many of the limitations of the coincidental method.

a. Compensation for coincidental telephone sample. The manner in which the size relations obtained in the roster method may be used to extend the coincidental measurement of actual audience size is found in a single mathematical principle: if the size of one object is known and the relation between the size of this object and the size of a second object is also known, the actual size of the second can be determined simply by multiplying the size of the first by the known size relation between the second and the first.

With the coincidental method, an exact measurement of audience size in one sub-group of the

total population—telephone homes in large cities—is obtained. This may, for example, be a rating of 10.0 and may represent the “first object”.

Simultaneously with the coincidental study, the roster method may be used to determine reasonably reliable relations of the audience size in any population sub-group to that in another. For example, the roster method may show the following relation: that telephone homes in large cities report listening to the given program in fifteen per cent of the cases, whereas a “cross-section” of all homes in small cities report listening to the program in twenty-five per cent of the cases. The relation between the audience sizes in those two sub-groups is, then, about twenty-five to fifteen or five to three.

Knowing that the size of the first sub-group’s audience (telephone homes of large cities) is 10.0, and knowing that the second is $5/3$ as great, the actual size of the audience in the second sub-group (total homes of small cities) is readily and accurately determined by multiplication.

Audience size in homes of small cities = $10 \times 5/3$
= 16.7.

This same simple process, involving the known audience size in one population sub-group and the size relation of that sub-group to any other can be used to determine, with reasonable accuracy, the audience size in any sub-group of the population or in the population as a whole. Thus, by combining the results obtained by the telephone coin-

cidental method with those of the personal interview printed roster, the limitations imposed on the coincidental by its telephone sample are overcome, and it becomes possible to measure reliably the audience size in all sub-groups of the total population. The roster method also compensates for the coincidental method's inability to yield data on flow of audience to and from a program and to measure duplication of listening between programs. The determination of audience size for network programs broadcast after 10:30 P.M. or before 8:00 A.M. is also made possible by this combination of methods.

The combination of the coincidental and roster methods has one notable advantage which is not found in some others. The roster portion can be set up at will at sufficiently low cost to permit the detailed study of a single advertiser's program and with no continuing overhead during periods of disinterest.

C. COMBINATION OF THE TELEPHONE COINCIDENTAL AND THE TELEPHONE DAY-PART RECALL METHOD

When a "computed" coincidental figure for telephone homes only is required the day-part recall method may be used in combination with the coincidental method, just as is the roster method. In fact, in the study of early morning programs, the coincidental-day-part recall yields a reasonably reli-

able computation which is somewhat simpler than those yielded by the coincidental-roster combination.

The method and its reliability is illustrated in a study made on the Pacific Coast in the cities of Los Angeles, San Francisco, Oakland, Seattle and Portland, all areas of "equal network opportunity". The regional ratings for each quarter-hour period between 7:00 and 8:00 A.M., local time, on each of the four major networks were required. Recall interviewing was conducted which covered the two-hour span, 7:00 to 9:00 A.M. Those programs broadcast between 8:00 and 9:00 A.M. were measured concurrently by the coincidental method. The regional rating for each fifteen-minute time period for each network was then computed as follows:

$$\frac{\text{Recall 7:00-7:15}}{\text{Recall 8:00-8:15}} = \frac{X}{\text{Coincidental 8:00-8:15}}$$

The rating for each quarter hour on each network was computed, using that network's 8:00-8:15 A.M. coincidental measurement as the base for computation. But since coincidental measurement was made on each of the quarter hours from 8:00 to 9:00 A.M., four separate bases for calculation were available. The stability of the computed figure could, therefore, be tested by computing the ratings on each of the four bases separately. The results obtained in using each quarter hour from 8:00 to 9:00 A.M. as the base are shown in Table XV.

TABLE XV

Stability of "Computed Coincidental" Ratings for
Early Morning Programs

(Based on Combined Coincidental Day-Part Recall Data
Obtained in Pacific Region)

Time Period	Network Coincidental Rating Used as Base	Computed Ratings			
		BNC	CBS	MBS	NBC
7:00-7:15 A.M.	8:00-8:15 A.M.	0.5	1.4	1.3	1.8
	8:15-8:30 "	0.4	1.4	1.4	1.5
	8:30-8:45 "	0.3	1.4	1.0	1.5
	8:45-9:00 "	0.6	1.4	1.3	1.7
	Average	0.5	1.4	1.3	1.6
7:15-7:30 A.M.	8:00-8:15 A.M.	1.0	1.6	0.6	1.3
	8:15-8:30 "	0.9	1.6	0.7	1.1
	8:30-8:45 "	0.8	1.6	0.5	1.1
	8:45-9:00 "	1.2	1.6	0.6	1.2
	Average	1.0	1.6	0.6	1.2
7:30-7:45 A.M.	8:00-8:15 A.M.	0.9	2.8	0.8	1.3
	8:15-8:30 "	0.8	2.9	0.9	1.1
	8:30-8:45 "	0.7	2.8	0.6	1.1
	8:45-9:00 "	1.1	2.9	0.9	1.2
	Average	0.9	2.9	0.8	1.2
7:45-8:00 A.M.	8:00-8:15 A.M.	0.5	2.9	0.8	4.0
	8:15-8:30 "	0.4	2.9	0.9	3.4
	8:30-8:45 "	0.4	2.9	0.6	3.4
	8:45-9:00 "	0.6	3.0	0.9	3.8
	Average	0.5	2.9	0.8	3.7

It was shown in Chapter VII that simultaneous study of programs by the coincidental and day-part recall methods is capable of yielding knowledge of

the factors influencing the memory for programs, which is of both practical and scientific value. From the 1941 studies in which the present writers compared the results obtained by these two methods, it was also possible to determine the "curve of forgetting" or the rate of memory loss for program listening.

The rate of loss during the two-hour period immediately following the broadcast is shown in Chart XL, which is based on a total of about 1,750,000 coincidental calls and the concurrent 7 months day-part recall sample. As would be expected, the loss increases rapidly with increasing length of the recall interval for periods up to an hour. After an hour, the loss increases much more slowly.

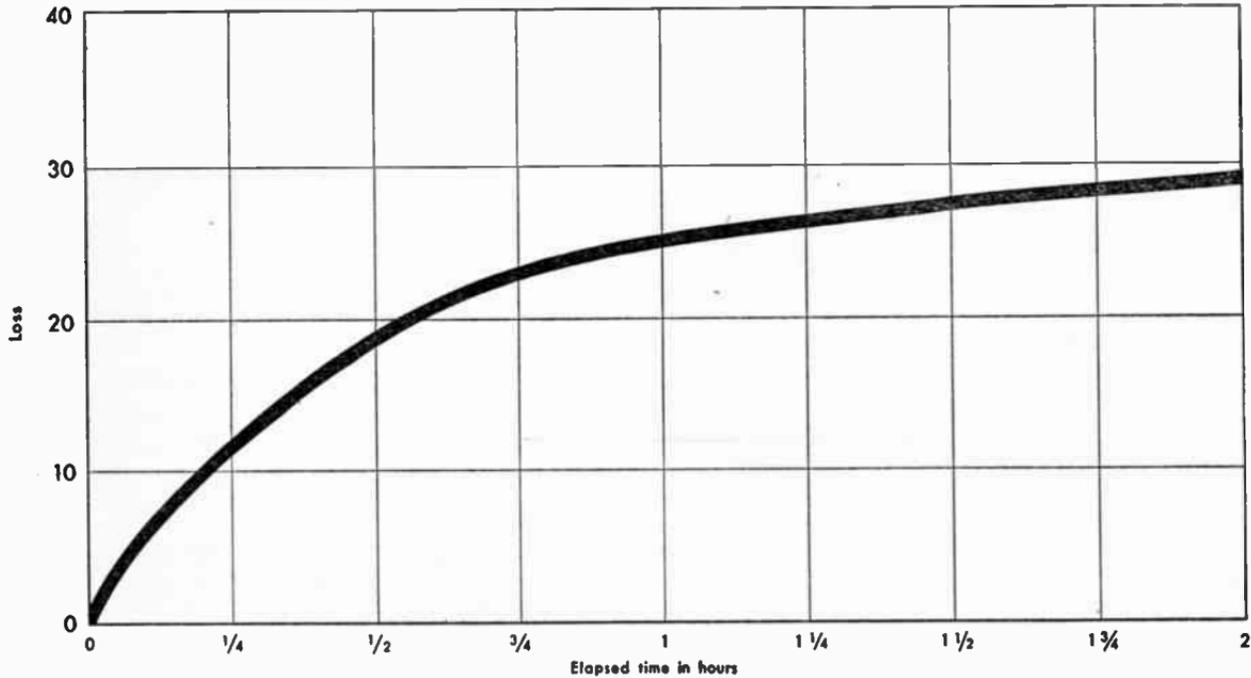
A. C. Neilsen, who has also studied this problem, reported results before The Market Research Council in New York in February, 1943, which might seem to controvert those represented in Chart XL. He compared results obtained by the mechanical recorder and the day-part recall methods and concluded that there was a 42 per cent memory loss in the first half-hour. Chart XL indicates a loss in the first half-hour of only 19 per cent.

A finding that broadcast material is subject to loss at so high a rate as 42 per cent in the first half-hour is somewhat astonishing in view of the mass of data accumulated on rates of forgetting by psychologists during the past fifty years. These data show a close relation between the ability to recall

CHART XL

MEMORY LOSS WITH ELAPSED TIME IN DAY-PART RECALL METHOD

Based on sectional coincidental and day-part recall ratings for 82 evening sponsored network programs. Each program rating was based on from 5 to 7 months' data.



and the degree of meaning of the material to be recalled. The most rapid rate of forgetting occurs in connection with material which is meaningless. This material is called "nonsense syllables" and is composed of combinations of letters such as wok, pam, jic, bip, seg, ron, taz, viz, lub, mer, koj, yad, etc., which have no meaning in the respondent's language.

Boreas, working with "nonsense material" in 1930, found at the end of 24 hours a loss of only 40.8 per cent. The rate of loss for names, numbers, words of the language, poetry and stories (of which broadcast material is composed) is much lower than the rate of loss for "nonsense" material. These findings, together with the lower rate of loss represented in Chart XL suggest that the 42 per cent difference between the mechanical recorder and the day-part recall results, which Neilsen attributed to memory loss, may contain both the memory loss and some other factor not recognized by the experimenter.

It was shown in Chapter VIII that the recorder is incapable of determining with exactness which short lines on the tape represent "votes for" a program and which "votes against". Further, Crossley's results indicate that the set is frequently in operation when there is no listening. Therefore, if all of the lines on the tape were assumed to represent listening, the recorder results would indicate more program listening than actually occurred. On failing to report programs to which the set was tuned but

which were not listened to, the respondent would be assumed to have forgotten. This, then, may be the source of the difference between Nielsen's 42 per cent and the writers' 19 per cent. If the 19 per cent represent accurately the actual loss in the first half-hour recall interval and the 42 per cent contains both this and some other factor, the magnitude of the other factor must be the difference between 42 per cent and 19 per cent or 23 per cent.

D. COMBINATIONS OF COINCIDENTAL AND FIXED SAMPLE METHODS

a. Coincidental and mechanical recorder methods. It was concluded in earlier chapters that the mechanical recorder method does not, and the panel method, if it comes into wide use, probably will not, yield valid measures of audience size. It seems probable, however, that these methods, like the printed roster, will yield reliable relations between the audience size in different population sub-groups.

Combination of the coincidental and the mechanical recorder methods would have about the same advantages and disadvantages as that of the coincidental and printed roster methods. The figures computed to determine actual audience size in the cross-section population would probably be somewhat more reliable for some programs but the cost would be much greater. The coincidental-mechanical recorder combination would yield data on avail-

able audience only in the coincidental sample. The same would be true of audience composition in terms of number, age, and sex of the listeners.

b. Coincidental and panel methods. The combination of the coincidental and panel methods seems to the writers to hold great potentialities. "Computed coincidental" ratings which would reveal the audience size in population sub-groups based on panel figures might prove to be quite as accurate as those based on the mechanical recorder. This combination would be expected to have all the advantages of the combinations discussed above and none of their limitations. Both audience composition in terms of number, age, and sex of listeners and available audience figures for all population sub-groups could be obtained by this combination.

5. Combination of Coincidental and Immediate Recall Methods

This combination was originally designed by the present writers to yield actual audience size measurements, but as was pointed out in Chapter VII it proved inadequate for that purpose. It did, however, prove to have outstanding merit for showing the flow of audience to and from programs.

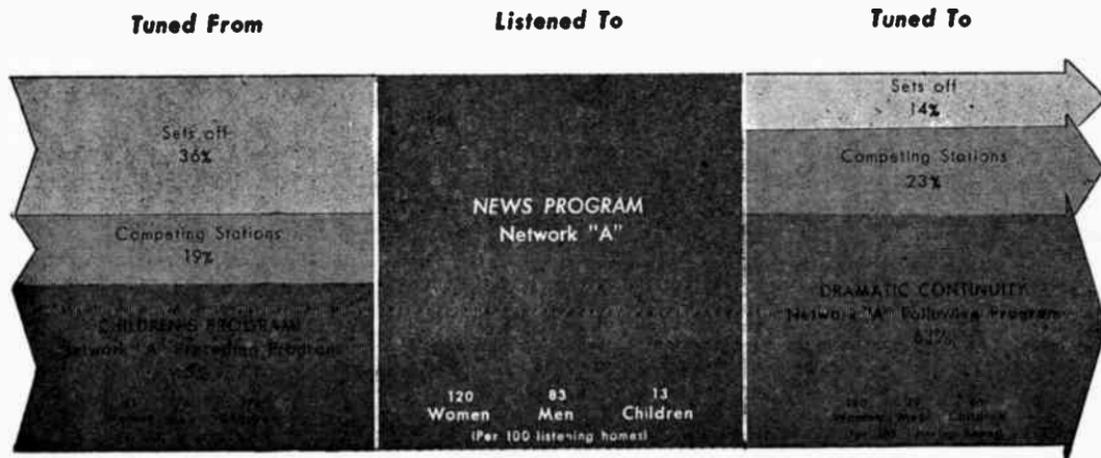
In most methods of measuring audience flow, e.g., the diary or the mechanical recorder, the record shows the audience flow only in terms of homes.

Nothing is revealed concerning the people who listened. For example, the recorder may show tuning to three consecutive programs in a given home. But there may have been no actual flow of audience; for the first may have been listened to by a child, the second by a man and the third by a woman. Measurements of audience flow achieve their full significance only when the composition of the flowing audience is known.

The combination of the coincidental and the immediate recall methods furnishes the audience flow in terms of people, not homes alone. In these studies, the respondent is asked both what he is listening to "now" and what he listened to 15 minutes earlier. If he is listening "now" he is asked how many men, women and children are listening. By conducting the interviewing continuously during consecutive 15-minute periods both the flow of audience and the composition are obtained.

Chart XLI illustrates the audience flow information developed from this combination of methods. The chart shows that 45 per cent of the homes which listened to the News program on Network "A" also listened to the Children's program which preceded it on that Network, but that the child audience disappeared precipitately when the News program came on. It is also noteworthy that more adults than children listen to this particular "children's" program. When the sponsor discovered this fact he revised his commercials immediately.

CHART XLI
FLOW "TO" AND "FROM"
A NATIONALLY SPONSORED NETWORK PROGRAM
Based on Combined Coincidental - Immediate Recall Data



E. COMBINATION OF "HOUSE-TO-HOUSE" AND TELEPHONE COINCIDENTAL METHODS

In the coincidental method, telephone interviewing has been used almost entirely. The use of the telephone as a means of interviewing is not essential in the coincidental method; but it is highly economical and, at the same time, has furnished much of the data required by the radio industry.

The limitations imposed on the method by using the telephone sample are largely overcome through the combination of telephone and personal interviewing techniques. No attempt is currently being made to establish a continuous combined coincidental operation covering a large area, because the method is so extravagant of manpower. But it is currently used for special studies made for individual clients. Its wider application, when manpower is available, is to be expected as a matter of course.

Very early in the experimental work with the personal interview coincidental method, the writers discovered that it was extremely difficult in the personal interview to obtain from telephone homes in the personal interview sample results which were strictly comparable with those obtained in the telephone coincidental sample. Analysis of the situation revealed that programs appealing to the higher socio-economic levels were rated higher in the telephone coincidental studies than they were in the telephone homes of "cross-section" sample. Questioning of the

interviewers revealed that, while the "cross-section" sample agreed with criteria such as the index to telephone ownership (per cent of families having residential phones), it was far from a true cross-section. Because so many families in the A and B economic levels live in apartment houses or other homes to which interviewers were denied access, the higher levels were grossly under-represented in the personal interview sample.

This stark fact of "cross-section" sample applies not only in obtaining a sample of radio listening homes but to the large city sample of personal interview study in general. It must be clearly recognized as a limit and limitation of personal interviewing in homes, and should give pause to those fond of extolling uncritically the accuracy of the personal interview approach.

Such a distortion of the sample may result in serious errors in results. It was found in the case of one local station having a high quality appeal that the listening to this station measured in the telephone homes of the personal interview coincidental sample was approximately twenty-five per cent less than was measured in the telephone coincidental study conducted concurrently. As was remarked earlier, all homes are reached with about equal ease by telephone. A way of obtaining a more perfect random sample of that part of the population in which buying power is concentrated would be difficult to conceive. The telephone subscriber

may live in a tower on a hill, well guarded by a doorman, or in a basement in a valley where his door is wide open; he may live in the center of town near three bus lines and four trolleys, or on the outskirts far from any public conveyance. He may have a vicious dog or he may not. Interviewers are human; therefore, all these factors determine whether or not a home is liable to be selected in a personal interview "cross-section" sample. They exert not the slightest influence on the telephone interviewer who is quite unaware of their existence.

Through the combination of telephone and personal interviewing techniques, the writers have developed a method for obtaining a proper "cross-section" of listening homes, one which represents properly homes inaccessible to personal interviewing.

Only a very small per cent of upper socio-economic level homes are found to be without telephones. For all practical purposes, it may be assumed that all upper level homes are equipped with telephones. Therefore, in order to obtain a "cross-section" of a city population, personal interviewing is conducted in the middle and lower socio-economic levels only. At the same time telephone interviewing is also conducted. The total population is represented by using from the personal interviews only those obtained in non-telephone homes and combining the results, properly weighted, with those obtained by concurrent telephone coincidental study.

The results for the lower and middle levels may be analyzed directly from the personal interview results alone. However, it is to be noted that there is no direct measurement of listening conditions in the upper level alone. The telephone coincidental results do not isolate the upper level. In a city such as Detroit with a telephone index of sixty-three per cent, all of the upper level homes, over half of the middle level homes, and a sizable percentage of lower level homes are equipped with telephones. The listening characteristics of the upper level in this procedure are, therefore, determined by subtracting from the results obtained for the total population the weighted results obtained for the middle and lower levels.

This combined personal and telephone interview coincidental method may be used either in the study of program or of station audiences. It is particularly useful in cases where a station introduces into the broadcasting situation by accident or design an element which offers a strong appeal to telephone homes only. For example, it may be found that station WAAA obtains only ten per cent of the listening in its local city, and it may obtain about the same audience in both telephone and non-telephone homes. In order to increase its listening the station may resort to some device for paying or otherwise rewarding "lucky" listeners. The device may operate in the following manner: Homes will be called on the telephone from time to time throughout the

day and the respondent asked what station he is listening to. If he says he is listening to Station WAAA and can identify the program being broadcast at that time, Station WAAA sends him a check for \$5.00.

If the reward is ample, such devices may attract to a station many listeners who would ordinarily listen to some other station. There is also some evidence to indicate that "reward" programs increase the amount of radio listening. That is to say, people listen to the reward-giving station when they otherwise would be listening to nothing. To the extent that the total radio listening is increased, "reward" programs may be considered to make some contribution to radio as a medium. However, station men who strive to build audiences with sound entertainment and service programs are prone to regard such programs as something of an abomination to the industry. But regardless of the attitudes of radio men or even the public toward stations that resort to these devices, they can, as in the case stated above, cause difficulty for audience measurers.

Since, in the example, Station WAAA uses the telephone to identify listeners, only those with telephones have any chance of winning the reward and only they can be motivated by the reward to greater WAAA listening. After some months of the practice, a station audience index based on telephone homes only might show that WAAA was now attracting a much larger per cent of listening. A com-

bined personal and telephone coincidental study of the city, which would include both telephone and non-telephone homes might show that, while WAAA had gained appreciably in telephone homes following the use of the "reward" device, the listening in non-telephone homes remained at its previous level or had declined.

This combination of interviewing techniques retains the fundamental strengths of the telephone coincidental method and eliminates two of the three limitations which apply to it. The combination is applicable to all populations and it yields results which lend themselves to analysis by the economic levels, etc. It still retains one limitation, though; it is not applicable to early morning and late evening programs.

The advantages of this combination are numerous. First, it is the only combination of methods which yields a measurement of the actual audience size in any and all population sub-groups. In all others discussed above the audience size must be computed. Secondly, it is the only combination which can measure the size of the available audience in any and all population sub-groups. Third, it is the only combination which measures the audience composition in terms of number, age, and sex in any and all population sub-groups. Fourth, it can be set up to measure a single advertiser's program or to measure continuously all programs, either local or network. And, finally, this combination is

relatively inexpensive; it costs less than a coincidental operation conducted by personal interviewing alone.

F. IN CONCLUSION

Methods and combinations of methods will come and go. Types of broadcasting will be perfected and supplanted by others. Even now frequency modulation and television are moving in where only amplitude modulation broadcasting was before. But one part of the broadcasting composite—the American public, without which neither radio nor its measurements would exist—will remain much as it is.

Rarely in this life do we get something for nothing. But in radio audience measurement and sampling research in general we have such a case. Radio audience measurements are available because the millions of people who make up America are willing to give to measurers the data they require. The greatest catastrophe that could be visited upon samplers would be the loss of the public's support and cooperation. One stupid act committed by an organization operating on a broad scale can threaten or even destroy all sampling operations.

If care is taken to ascertain that each individual is called upon at infrequent intervals, and if each is approached with the courtesy and consideration due one who is granting a favor, public attitudes in line with the continued good health of sampling operations in general will obtain.

End

GLOSSARY

Advertising Media: Vehicles for carrying advertising, such as radio, newspapers, magazines, billboards, car cards, direct mail, etc.

Areas of Equal Network Opportunity: Those areas in which all four major networks can be heard with equal ease and are habitually heard; e.g., within cities in which each of the four major networks have a local outlet.

Attenuation: The earth's ability to absorb the station's signal. This is not a constant but varies area by area and according to the station's frequency.

Audience Composition: The number of men, of women, and of children per listening set found listening to a given program or listening during a given time period.

Audience Turn-over: See "Frequency of Listening".

Available Audience: Per cent of total homes sampled in which someone is "at home and awake".

Average Audience: Audience measurement obtained when listeners to the entire program, and listeners to parts of the program receive the weight which the length of their listening time dictates.

Average Listening Time: Average number of hours, minutes or seconds of listening during a specified period of time.

Basic Network Area: The area served by basic network stations, being different for each network.

Basic Network Station: A station, in a group forming the network nucleus, which the advertiser is usually required to include in his minimum network purchase.

Broadcaster: The station or the network transmitting the program.

Call Letters: The combination of three or more letters or

numbers assigned by the Federal Communications Commission which constitutes a station's official designation. By international law, United States stations carry N, W or K as the first letter.

Continuous Measurements: Those made all the time or at regular, frequent intervals which permit a development of trends.

Coincidental Method: The method of measuring the radio audience while it exists—during the broadcast.

Commercial Announcement: Any radio advertising message. The following are classes of commercial announcements.

1. *Program Commercials:* advertising messages carried as a part of a specific program identifying the advertiser or product responsible for the broadcast of that program. Program commercials may be of three kinds:
 - a. *Opening:* located in the first few minutes of program time.
 - b. *Middle:* set in the entertainment. These may be:
 - i: *Program break:* which requires an interruption of the entertainment while the commercial is presented; or
 - ii: *Integrated:* a part of the program entertainment requiring no interruption.
 - c. *Closing:* located in the last few minutes of program time.
2. *Cow-catchers:* advertising messages carried just before the specific program and advertising some auxiliary product of the advertiser presenting the program which follows it.
3. *"Trailers"; or "Hitch-hikes":* advertising messages carried just after the close of a specific program and advertising some auxiliary product of the advertiser presenting the program which it follows. Frequently it follows immediately the closing program commercial.
4. *"Chainbreak":* any commercial placed in the station identification period. It may be as short as 20 words or as long as 20 seconds. It is presented by an advertiser

who presents neither the preceding nor the following program.

5. "*Cut-in*": a local advertising message presented by the local station during an interruption of a network program.

Conscious Impression: An impression which can be recalled subsequently; usually measured in terms of memory for name of sponsor, program, product or station.

Controlled Experiment: One in which the influence of all variables except that one studied are eliminated or shown not to operate significantly.

Correlation: A statistical term showing the degree of concomitant variation between two factors or the extent to which variation in one must be attended by variation in the other.

Coverage Area: Originally an engineering term measured physically; but in current use, the area, usually measured in countries, in which a station can be and is habitually heard. No criteria for "habitually heard" are yet generally accepted.

Crossley: A term used generically in the radio industry to indicate a program rating. It derives from Archibald M. Crossley, the oldest name in radio audience measurement.

Day-Part Recall Method: An unaided memory method in which the respondent is asked to report on previous listening during a specified span of hours.

Diary Method: A panel method designed to study radio audiences for short periods of time, usually one week.

D.K.: Letters used in all sampling survey work to indicate a "don't know" response in answer to some question.

Directional "Broadcasting" Transmission: A radio antennae system which concentrates the signal in a desired direction.

Duplication of Audience: The per cent of the people reached by one program which also listens to another program.

Fixed Sample: One in which data are obtained repeatedly from the same sources—either the same people, the same machines, or the same brick houses.

Flow of Audience: Measurements showing where a program's

audience comes from and where it goes after the program ends. Complete measurements report the amount and composition of the flowing audience.

Frequency of Listening: In the case of a five-a-week program, the average number of times the program is heard in one week. In the case of a one-a-week program, the average number of times the program is heard in four weeks.

This term is closely related to "Audience Turnover" which is computed by dividing the number of broadcasts on which "frequency" is determined by the average frequency.

Ground Wave: That part of the radio signal which travels from the radio station antenna to the receiver in a plane horizontal to the earth. This signal is gradually attenuated or absorbed as it travels from the transmitter.

Hooperating: A term originally used generically in the radio industry to indicate all kinds of radio ratings whether on programs or time periods. It was subsequently adopted by C. E. Hooper, Inc., to identify its rating services based on the coincidental method. Thus, the Hooper report containing ratings on network programs is named "Network Hooperatings Report", etc.

Immediate Recall Method: An unaided memory method in which the respondent is asked what program he listened to up to 15 minutes before the interview.

Independent or Non-network Station: One having no major network affiliation.

Inherited Audience: That part of a program's audience which listened to the program which preceded it on the same station or network.

Interference: Undesirable electrical noises affecting radio reception. These may be man-made (x-ray machines, other radio stations, etc.) or natural (static).

Internal Pattern of Listening: Program listening analysed in terms of ratings for successive 1, 3, or 5 minute intervals of program time.

Local Outlet: A station broadcasting from within a community.

Mail Count: The number of pieces of mail attracted by a

program, a specific broadcast, a campaign, or a station. The mail may be unsolicited, solicited without reward, a direct mail order for merchandise, the response to a contest, or prompted by premiums.

Mail Map: A map designed to show a station's listening area which is plotted from data obtained through analysis of amount, frequency and local origin of station mail.

Major Network: A network organization and the stations affiliated with it to form a transcontinental broadcasting service, currently: Blue Network Company, Columbia Broadcasting System, Mutual Broadcasting System, and National Broadcasting Company.

Mean: An average.

Mechanical Recorder: A method in which listening characteristics are inferred from lines made on a tape by a mechanism attached to the radio set.

Median: The midpoint of a number of measurements. Thus with 15 ratings, the 8th from the top or the bottom would be the median. With 16 ratings (or any even number) there is no single figure which represents the midpoint. In this case, the median is the value which lies halfway between the 8th and the 9th rating.

Mental Set: All those mental characteristics—habits, general interests, momentary interests, attitudes, emotions and mental activity in progress which constitute one's frame of mind at any given time.

Modulation: The process of altering a carrier wave so that intelligence (words, music, telegraphic code, etc.) is transmitted by it.

a. **Amplitude Modulation (AM):** a system of modulation whereby the frequency of the carrier wave is held constant and the power is varied in accordance with the intelligence to be transmitted.

b. **Frequency Modulation (FM):** a system of modulation whereby the power of the carrier wave is held constant and the frequency is varied in accordance with the intelligence to be transmitted.

Moving Average: An average based on the two or more con-

secutive checks in a series of sampling operations. Commonly used where the sample for each individual study is too small to yield reliable figures.

Network Affiliate: A station in contract with and transmitting sponsored or sustaining programs from a network.

Network Competition: Programs carried on other networks while the "subject" program is being broadcast.

Network Option Time: Specific hours on a station affiliated with a network which, on request, the station has contracted to clear, within 56 days, for broadcasting network programs.

Noise Level: The level or strength of interfering signals.

"One Shot" Measurements: Those taken once and not repeated to permit development of trends.

Panel Method: A method employing a "fixed" sample in which respondents keep a record of all activities in certain categories such as radio listening and brand purchases.

Product Use Study: A sampling study made to determine the use of a sponsor's brand of product and that of his competitors among listeners to his program compared with non-listeners—typically revealing influence of both length and frequency of listenership on degree of use.

Program Effectiveness: The extent to which a program produces results desired by the sponsor.

Program-part Listeners: People who listen to only part of a program.

Program Type: A program's classification according to outstanding characteristics of content or form; e.g., news, variety, serial, quiz, popular music, etc.

Projection: The calculation of the number of people in a population listening to a program through the use of a program rating (based on homes listening) and a measure of the number of program listeners per listening home. Ratings are projected only to the population represented in the sample from which the rating is obtained.

Random Sample: One in which all members of a population have equal opportunity for being selected as members of the sample.

Ratings: In general, the percent of a sample found listening to a specific program, station or network.

Coincidental Rating: A measurement in which the base, 100%, is total homes called, including those in which no one is at home. The rating is the per cent of the total calls in which the respondent reports that he is listening to a given program being broadcast at that time.

Mathematically, a coincidental rating equals:

$$\frac{\text{Total calls—Don't Answers}}{\text{Total Calls}} \times \frac{\text{Yes}}{\text{Yes} + \text{No to Radio}} \quad (\text{Listening to Radio})$$

$$\times \frac{\text{Listeners to Programs}}{\text{Yes (Listening to Radio) — D.K. Programs}}$$

Conscious Impression Rating: A measurement in which the base, 100%, is the total persons questioned and the rating is the per cent of those persons who recall the name, identifiable talent or the sponsor of a program previously listened to within a designated span of hours, days or weeks.

Cross-section Program Ratings: Measurements which reveal the total number of families listening in the national population or in a large geographical segment of it; e.g., total homes listening in the nation, in the Eastern Time Zone, or in cities having a population of 25,000 and over. It is differentiated from a "network program rating" in that its sample must represent reliably all variations in coverage and all combinations of station competition whereas these two variables must be completely controlled or eliminated in obtaining the "network program rating" which measures program attraction value.

Day-part Recall Rating: A measurement in which the base, 100%, is total homes in which someone is at home at the time of the interviewer's call and the rating is that per cent of this group who, without aid, remember having listened to a program previously broadcast. The time span on which the respondents report is usually the two hours immediately preceding the interviewer's call.

Limited Network Program Rating: Defined by C. E.

Hooper, Inc., as a rating obtained on a program broadcast in fewer than 15 of the 32 standard Hooper checking cities and/or fewer than 3 of the 5 Hooper geographic areas.

Mechanical Recorder Rating: A measurement in which the base, 100%, is total homes equipped with recorders and the rating is the per cent of these homes in which the radio is in operation and tuned to a given station.

Panel or Diary Rating: A measurement in which the base, 100%, is the homes of the pre-selected "fixed" sample which submit a report. The rating is the per cent of those reporting who report name, identifiable talent, or sponsor of a program listened to within a designated span of time.

Roster Rating: A measurement in which the base, 100%, is total homes in which someone is at home at the time of the interviewer's call and the rating is the per cent of this group which, on being presented with a list of programs, reports listening to the specific broadcast.

Sectional Rating: A rating obtained in one geographic area.

Time Period Rating: A rating covering a given span of time irrespective of program or programs carried.

Uniform Network Competition Rating: Per cent of the sample listening to a program in those cities where the programs carried concurrently by one or more other networks are the same for all cities.

Recall Interval: Length of time between the end of a broadcast and the subsequent recall interview in a "recall" study.

Recruiting Efficiency: The per cent of the available audience listening to the radio (as opposed to "sets-in-use", which is the per cent of the total sample listening to the radio).

Representative Sample: One which reflects accurately all the states of mind present in the population which are important in determining the characteristic to be measured.

Respondent: One who replies to questions in a sampling operation.

Reward Program: A local program which gives prizes or money to contacted listeners if they report listening to that program.

Roster Method: An aided memory method in which the respondent is presented with a list of programs and asked to report on his listening.

Sampling Operation: One through which specific characteristics of a population are determined by examination of a small segment of that population.

Sets-in-Use: Per cent of the total sample found listening to the radio. Mathematically it equals:

$$\frac{\text{Total calls—Don't answers}}{\text{Total Calls}} \times \frac{\text{Yes}}{\text{Yes} + \text{No to Radio}} \text{ (Listening)}$$

Share of Audience or % of Listeners: The per cent of all listeners which is found listening to a specific program, station or network at a given time. Mathematically it equals:

$$\frac{\text{Listeners to Program}}{\text{Total Listeners}}$$

Listening to radio — Listening but D.K. Program

Short Wave: Waves having lengths of from 10 to 100 meters, now correctly referred to as high frequencies.

Show: Synonymous with program.

Signal Strength: The electrical magnitude of a station's radiated signal usually stated in millivolts (thousandths of a volt).

Skywave: That part of the station's signal which travels from the antenna to the sky and then is reflected back to the earth.

Sponsor Identification Index: The per cent of listeners to a specific program which knows the name of the program's advertiser, or of any of his products. The sponsor identification analyses breaks down the S.I.I. into Product Advertised on Program and other correct identifications and in addition reports misidentification and Don't Know.

Sponsored Broadcast Hours: The sum of the time periods devoted to the broadcast of nationally sponsored network programs, expressed in hours.

Sponsored Program: One presented during time for which the broadcaster is paid.

Spot Broadcast: A non-network program of a national advertiser.

Standard Broadcast Band: The range of frequencies from 550 to 1600 kilocycles.

Station Competition: Programs carried on other stations while the "subject" program is being broadcast.

Station Frequency: Number of waves per second set up by the station's transmitter, usually given in terms of kilocycles (or thousands of cycles). Generally known as the station's spot on the dial.

Station Listening Index: The proportion of listening to each station heard in a given city reported by portions of the broadcast day.

Station Identification Period: The time interval in which a station announces its official call letters as required by the Federal Communications Commission.

Station Power: The energy rating of the transmitter given in watts or kilowatts.

Station Service Area or Listening Area: The area over which a station is listened to habitually in a substantial number of homes.

Statistical Error: The variation inherent in all sampling operations. The size depends upon the size of the sample. The amount of the error decreases as the square-root of the increase in sample size.

Statistical Reliability: In general, the degree to which a figure obtained by sampling approximates a true measure. The following are measures of statistical reliability:

(a) **"Indicative" vs. "Conclusive" Ratings:** A device developed by the writers to engender some caution in the use of ratings. An "Indicative" rating is one based on less than 600 calls and is, therefore, subject to appreciable statistical variation. A "Conclusive" rating is one based on 600 or more calls and has smaller limits of variation. Whether "Indicative" or "Conclusive", the statistical tolerance of all ratings should be determined (See Chart II) before basing decisions on them.

(b) **Probable Error:** One half of the range which theoretically would contain the middle 50 of 100 measurements made on the same population using the same

sample size. It has little accurate significance to those not statistically trained.

(c) *Standard Deviation or Sigma or Standard Error*: One half of the range which, theoretically, would contain the middle 68 of 100 similar measurements made on the same population using the same sample size. It has little accurate significance to those not statistically trained.

(d) *Statistical Tolerance, Significant Difference, or Maximum Expected Variation*: All three of these terms are used to indicate the amount by which one measurement must be larger or smaller than another obtained from a similar sample in order to indicate an actual difference. The term "statistical tolerance" is probably original with the present writers.

Statistical Validity: The degree to which a sampling figure represents conditions that actually exist.

Stratified Sample: One selected in accordance with known population characteristics and developed for internal consistency of all its parts.

Sustaining Program: One presented during time paid for by the station or network.

Television: A system of transmitting instantaneously combined sound and moving visual images.

Time:

E.T.: Eastern Time

C.T.: Central Time

M.T.: Mountain Time

P.T.: Pacific Time

Total Homes Listening: The number of homes in an area or in the nation listening.

"Verified" Listeners and "Verified" Non-Listeners: Homes in which listening and non-listening is double checked. (Used in connection with product use studies made by C. E. Hooper, Inc.).

Video: An adjective relating to signals employed in the visual portion of the television system.

Weighting: A mathematical adjustment of sampling results which compensates for a known non-representative distribution of a sample.

INDEX

- "Advertising & Selling", 175 f.
 Allen, Fred, 18
 Ameche, Don, 84
 American Association of Advertising Agencies, AAAA, 188
 Association of National Advertisers, ANA, 4, 188
 Archer, 1
 At Home (*see* Available Audience)
 Attenuation, glossary
 Audience,
 Available —, 12, 62, 81, 198 ff., 219, glossary
 Average —, 67 f., glossary
 — Composition, 88 ff., 172, 199, 219, 227 f., glossary
 Duplication of —, 197, 199, 204 f., 209, glossary
 Flow of —, 171, 191, 196, 199, 209, 219 ff., glossary
 Inherited —, 82 ff., glossary
 — Turnover, glossary
 Autrey, Gene, 84
 Average, Moving, glossary (*see* Audience)
 Ayer, N. W. & Son, 207

 Barton, Samuel, 198
 Benny, Jack, 14
 Blankenship, Dr. Albert, 206
 Blondie, 86, 76
 Blue Network Company, BNC, 18 f., 17 f., 22, 87, 92 f., 147 f.
 Boreas, 217
 Bridgman, 6
 Broadcast,
 Standard — Band, glossary
 Spot —, glossary
 Variation in —, 11 f.
 Broadcaster, glossary
 Broadcasting,
 Directional Transmission —, glossary
 — Variables, 11
 — Conditions, 14
 Burns & Allen, 18

 Cavalcade of America, 198
 Cawl, Franklin R., 189, 198
 Chappell, M. N., 189
 Churchill, John; acknowledgment, 198
 Clark-Hooper, Inc., 4
 Coca-Cola, 8
 Coin Tossing, 5 ff.
 Coincidental (*see* Methods)
 Columbia Broadcasting System, CBS, 18 f., 22, 92 f., 198, 205
 Commercial, Announcement, 82 f., 200, glossary
 Chainbreak, glossary
 Closing, glossary
 Cow-catcher, glossary
 Cut-in, glossary
 Hitch Hikes, glossary
 Middle, glossary
 Opening, glossary
 Trailers, glossary
 Conscious Impression (*see* Impression)
 Cooperative Analysis of Broadcasting, CAB, 4, 16, 61, 188 f.
 Correlation, glossary

- Counterspy, 18
 Coverage, 12 f., 21 f., 192 f., 197
 Crosby, Bing, 18
 Crossley, Archibald M., Acknowledgment, 4, 188, 162, 175 f., 182, 207, 217, glossary
- Data,
 Influence of —, 45
 Obsolete —, 45
 Deviation, Standard —, 50, glossary
 Difference,
 Significant —, 50 ff., glossary
 Reliable —, 55 ff.
 D.K., glossary
 DuPont, 198
 Easy Aces, 86 f.
 Economic Status, 89, 71, 165, 171, 209, 222 ff.
 Eddy, Nelson, 84
 Educational Status, 87, 209
 Effectiveness, Sales, 191 (*see* "Program")
 Error,
 Probable —, 50, glossary
 Sources of —, 45
 Standard —, 50, glossary
 Statistical —, 59 (*see* "Statistical")
 Family Composition, 85, 208
 Famous Jury Trials, 18
 Fibber McGee & Molly, 14, 80 f., 85 ff., 185 ff.
 Fitch Bandwagon, 18 f.
 Freiberg, Albert, 7, 8
 Frequency,
 Station —, 22, glossary
 — of Listening, 68, glossary
 — modulation, 228, glossary
 Gallup, George, acknowledgment, 206
 Gangbusters, 18
 Geographic,
 — Areas, 158 ff., 172
 — Conditions, 84
 Gill, Samuel E., Acknowledgment, 185 f.
 Heatter, Gabriel, 18
 Hooper, C. E., Inc., 4, 5 ff., 87, 61 f., 69 f., 78 ff., 189, 207
 Hooperating, glossary
 Hope, Bob, 14, 85 ff.
 "Horse & Cart", problem, 108 f.
 Hour of Charm, 77
 Impression, Conscious, 101 ff. (*see* Rating)
 Index,
 Available Audience —, 112 ff.
 Average Rating —, 112 ff.
 Sets-in-Use —, 112 ff.
 Sponsored Broadcast Hours —, 112 ff.
 Station Listening —, 73, 74, glossary (*See* Trends)
 Industrial Survey Company, 198
 Intelligence, 87
 Jergens Journal, 18
 Johnson's Wax, 8
 KSL, 105
 KLZ, 105
 Lamb, Charles, Introduction
 Laws, scientific —, Introduction
 Lazarsfeld, Paul, 89
 Lin Yutang, 49
 Listeners,
 Non-Verified —, 96 ff., glossary
 Program-part —, 67, 162 f., glossary
 Total —, 67
 Verified —, 96 ff., glossary
 Listening,
 Conditions influencing —, 28 f.
 Continuing Measurement of —, 106
 Duplication of —, 172, 197, 199, 209

Listening—Continued

- Frequency of —, 172, 199, glossary
- Internal pattern of —, 82 f., 171, glossary
- Station —, 78 f., 105
- Valid measure of —, 64, 178 ff.
- "Literary Digest," 206
- Locality, Size of, 84
- Lone Ranger, 85 f.
- Lux Radio Theater, 11 f., 92, 105
- Mall,
 - Count, 2 f., glossary
 - Map, glossary
- Market Research Council, 215
- Major Bowes, 11, 80 f., 66
- Manhattan Merry-Go-Round, 27
- Mayor of the Town, 11
- McCarthy, Charlie, 14, 88
- Mean, glossary
- Measurement,
 - Continuous —, 106, glossary
 - of specific instances, Introduction
 - "One Shot", —, 82, glossary
 - Units of —, 7, 66 f.
- Media, Advertising, glossary
- Median, glossary
- Memory loss, 159, 215 ff. (*see* Variables)
- Mental set, 26 f., 29
- Mental tuning out, 101 f.
- Methods,
 - Applicability of —, 68 ff., 184, 209 ff.
 - Call Back —, 97
 - Case History — (*see* "Panel")
 - Coincidental, House to House —, 222 ff.
 - Coincidental, Telephone —, 61 ff., 111 f., 140 ff., 209 ff., glossary
 - Combinations of —, 206 ff.
 - Consumer Jury — (*see* "Panel")
 - Day-part recall —, 60 f., 188 ff., 207 f., 212, glossary
 - Diary —, 202, glossary
 - Fixed sample —, 178 ff.
 - Immediate Recall —, 168 ff., 219 f., glossary
 - Limitations of —, 228
 - Mechanical Recorder —, 5, 60, 81, 207, 218 f., glossary
 - Panel —, 197 ff., 219, glossary
 - Recall —, 188 ff., 201
 - Roster —, 60 f., 168 ff., 218, glossary
 - Mutual Broadcasting System, MBS, 18, 22, 147 f.
 - Mystery Theater, 18
 - National Broadcasting Company, NBC, 18 f., 22, 85, 92 f., 147 f.
 - Network,
 - Affiliation, 24, glossary
 - Basic — area, glossary
 - Competition, 8, glossary
 - Equal — opportunity, 10, 19 f., 70, glossary
 - Major —, glossary
 - Option Time, glossary
 - Outlets, 22, 70
 - Uniform — Competition, glossary
 - News of the World, 190
 - Nielsen, A. C., Company, 5, 78, 107, 170, 171, 185, 217
 - Noise Level, 21, glossary
 - Non-Listeners,
 - non-verified —, 96 ff., glossary
 - Verified —, 96 ff., glossary
 - "Not-at-Home", 149 ff. (*see* Sample)
 - Occupation, 88, 172, 209
 - Old Fashioned Revival, 27
 - Opportunity, Areas of equal, 70 (*see* Network), glossary
 - Origin, National, 88, 172
 - Outlet,
 - Local —, 24, glossary
 - Non-Network —, 24, glossary (*See* Network)

- Parmenides, 188
 Pearl Harbor, 119 ff.
 Phillip Morris Play House, 18
 Population, 10 ff., glossary
 Sub-groups of —, 43 f., 80 ff.,
 227
 Pot o' Gold, 84
 Preference,
 Brand —, 96 ff., glossary
 Program —, 7, 64 f., 75 ff., 179 ff.
 Principles, Fundamental, 7, 109
 f., 172
 Product Use, 97 ff., 199, glossary
 Program,
 Age of —, 141
 Analysis of —, 191
 Availability of —, 29
 Children's —, 85 f., 91
 Effectiveness of —, 96 ff., 104 ff.,
 190 f.
 Length of —, 141
 Popularity of —, 144
 — Preferences (*see* Preference)
 Reward —, 225 ff., glossary
 Sponsored —, 144, glossary
 Sustaining —, glossary
 — Type, 144, 146, glossary
 Projection, glossary
 Psychological Corporation, 2

 Quaker Oats, 8

 Radio Corporation of America,
 RCA, 1
 Radiograph Corporation, 170, 175
 Radio Readers Digest, 27
 Ratings, 6
 City-by-City —, 95 f.
 Coincidental —, 62 ff., 108 ff.,
 116 ff. (*see* "Methods"), glos-
 sary
 Comparability of —, 16
 Computed —, 211 ff., 219, glos-
 sary
 Conclusive —, 79, glossary
 Conscious Impression —, glos-
 sary
 Cross-Section —, 16 ff., 62, glos-
 sary
 Day-Part Recall —, 189 ff., 211
 (*see* "Methods"), glossary
 Diary —, 202 ff., glossary
 Immediate Recall —, 168, 219,
 glossary
 Indicative —, 79, 96, 106 f.,
 glossary
 Limited Network program —,
 glossary
 Mechanical Recorder —, 195,
 196, glossary
 Network Program —, 15 ff., 62,
 95, glossary
 Panel —, 197 ff., 218 f., glossary
 Roster —, 168 ff., glossary
 Sectional —, 156, 160, glossary
 Time Period —, 107, glossary
 Recall,
 Day-Part
 — Interval, 157 ff., 215 ff., glos-
 sary
 — Rating (*see* "Ratings")
 Twenty-four Hour —, 4, 188
 Recruiting Efficiency, 62, 79,
 glossary
 Research,
 Advertising —, Introduction
 Media —, Introduction
 Psychological —, Introduction,
 215 ff.
 Respondent, 64 ff., glossary
 Robinson, Ray, 197
 Roosevelt, F. D., President, 18
 Roth Orchestra, 86

 Sales Effectiveness (*see* "Program
 Effectiveness")
 Sample,
 "At Home" —, 149 ff.
 Errors in Stratification of —,
 45 f.
 Fixed —, 188, glossary
 — ing Operation, 25 ff., glossary
 Minimum —, 96, 105
 "Not at Home" —, 149 ff.

- Sample—*Continued*
 “Not at Home Then” —, 157 ff.
 Random —, 40 ff., 45, 47 f.
 Representative —, 25, 69, 188, glossary
 — Size, 78 ff., 189 ff.
 Stratified —, 42 ff., 172, 185 ff., glossary
 Telephone —, 46 f., 71 ff., 209 ff., 222 ff.
- Science,
 — of Mass Behavior, Introduction
 — of Social —, Introduction
 Screen Guild Theater, 14
 Sets-in-Use, 68, 94, glossary
 Share of Audience, 184, 186 f., glossary
 Show, glossary
 Sigma, 59, glossary (*see* “Statistical”)
 Signal Strength, 21, glossary
 Smith, Charles, Acknowledgment, 198
- Sponsor,
 — ed Broadcast Hours, 2, 115 f., 122 f., 129 f., 184 f.
 — identification, 102 f., glossary
- Spot Light Bands, 165
- Stanton, Frank, Acknowledgment, 198
- Station,
 Basic Network —, glossary
 — Call Letters, glossary
 — Competition, 22, 24, 82 f., 187, glossary
 — Coverage, 192, glossary
 — Identification Period, glossary
 Independent —, 24, glossary
 — Listening, 78, 105 f.
 Local —, 77
 Non-network —, 24, glossary
 — Service Area, 106 ff., glossary
 — Signal Strength, 21, glossary
- Statistical (*see* “Variations”)
 — Error, 50 f., glossary
 — Reliability, glossary
 — Significance, 50 f., 59, 78 ff.
 — Tolerance, 57 ff., glossary
 — Validity, glossary
 — Weighting, glossary
- Statistics, 49
- Summers, Leda P., 87
- Telephone Hour, 18
- Telephone Sample (*see* “Sample”)
- Television, 228, glossary
- The Thin Man, 67
- The Unseen Audience, 174
- This Is War, 83
- Time,
 — Listening —, 171, glossary
 — Zones, 171
 — Tuning —, 175 f.
- Tolerance (*see* “Statistical”)
- Trends,
 Basic Industry —, 80, 111 f., 117 ff.
 Program Type —, 181 ff.
 (*See* “Indexes”)
- Tuning,
 Total Homes —, 90 f.
 — Time (*see* “Time”)
 Types of —, 176 ff.
- United States Census, 41 f., 46
- Variables,
 Broadcast —, 11 f.
 Memory —, 140 ff.
 (*See* “Statistical”)
 — Units of Measurement, 181 f.
- Variation,
 Chance —, 51, 55 ff.
 Geographic —, 158 ff., 172
 Maximum Expected — (*see* “Statistical Tolerance”)
 Seasonal, 149
 (*See* “Variables”)
- Video, glossary
- Waltz Time, 18

- Wave,
 Ground —, glossary
 — Length, glossary
 Sky —, glossary
Weighting (*see* "Statistical")
Winchell, Walter, 18, 26 f.
- WLW, 74, 107 f.
"Woman's Home Companion",
 197
Woodworth, Professor Robert S.,
 26
Wylie, Max, 49

