THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 50, No. 5, May-June, 1971
Printed in U.S.A.

Listener Evaluation of Simulated Telephone Calling Signals

By P. D. BRICKER

(Manuscript received December 2, 1970)

This research concerns the judged pleasantness of a variety of electronic calling signals. Five experiments are reported in which type and frequency of carrier; type, frequency, and waveform of modulation; and spectral composition were varied. The results have aided in the selection of two signals for further trials.

I. INTRODUCTION

Technological considerations suggest that an electronic "ringer" may succeed electromechanical devices in future telephones. This possibility has generated interest within the telephone industry in the specification of desirable characteristics for electronic calling signals. Not the least important of these characteristics is that such signals be acceptable to the subscriber on purely aesthetic grounds. The literature on tone ringers includes some references[1,2] to the measurement of listeners' opinions, but only recently has any systematic work on what constitutes a pleasant signal begun to appear.[3] An earlier study by P. D. Bricker and J. L. Flanagan[4] marked the beginning of an attempt to chart the preference-relevant dimensions of a fairly large class of calling signals. The present paper reports five subsequent experiments that have clarified the effects on evaluative judgments of a half-dozen parameters. These experiments have interacted with studies of calling-signal detectability and with development work to produce two distinct realizable signals, which are scheduled for evaluation in a field trial.

II. EXPERIMENT 1

2.1 *Background*

The first experiment was identical in form to that reported by Bricker and Flanagan.[4] That is, listeners heard one signal at a time and assigned each a number that reflected their opinion of the signal. This technique, in conjunction with a means of analyzing the data in terms of perceptual attributes, had provided considerable information about a limited variety of signals in the earlier study. The purpose of the present experiment was to obtain a rough idea of the parameters important to evaluation in a much larger domain of signals, to serve as a guide to more detailed investigations.

2.2 *Signals*

There were 100 different signals in this experiment of three basic types:

(*i*) Thirty-four *a*mplitude-*m*odulated *p*ulse-train-*c*arrier (AMPC) signals, which were a subset of the signals studied by Bricker and Flanagan.[4] These represented selected combinations of four modulation frequencies, three duty factors, three carrier frequencies, and three harmonic compositions.

(*ii*) Six *a*mplitude-*m*odulated *s*inusoidal-*c*arrier (AMSC) signals, representing three modulation frequencies and two carrier frequencies.

(*iii*) Sixty *f*requency-*m*odulated *s*inusoidal-*c*arrier (FMSC) signals, representing selected combinations of six modulation frequencies, three carrier frequencies, three amounts of frequency deviation, and five modulation waveforms.

2.3 *Listeners*

Forty-three persons of various nonsupervisory employment classifications at Bell Laboratories served as listeners.

2.4 *Procedure*

Groups of four to six listeners, seated around a table in a carpeted room with draperies, listened to one of four permutations of the 100 signals reproduced over a high-quality magnetic tape playback system. They were instructed to record a positive number on the answer sheet for signals they liked and a negative number for signals they disliked; the greater the degree of liking or disliking, the larger the positive or negative number. A new signal occurred every 6 seconds, so that the entire procedure, including reading the instructions and rest periods, required about 15 minutes.

2.5 *Analysis*

The listeners' ratings were arranged in a matrix of 43 rows (listeners) by 100 columns (signals) and each row was normalized so that it had $\mu = 0$ and $\sigma = 1$. These data were analyzed by the MDPREF computer program of J. J. V. Chang and J. D. Carroll[5] so as to produce a spatial representation of both signals and listeners. MDPREF solutions represent the stimuli (calling signals, in this case) as points and the subjects (listeners) as vectors in multidimensional space in such a way that the projections of the points on each subject's vector correspond maximally, in a least squares sense, to his input data vector. Another property of these solutions is that successive dimensions are orthogonal to those preceding and account for as much of the residual variance as possible. It is left to the experimenter to determine how many solution dimensions will be considered significant and how he will rotate the axes to render the solution interpretable.

Another technique found useful in interpreting the results of this experiment was to regress the coordinates of the stimulus points on physical property vectors, so as to locate vectors maximally corresponding to the parameters that were varied to generate the stimuli. In the earlier experiment,[4] regression techniques had been used to find a three-dimensional structure in the data that was interpretable in terms of three parameters of signal design.

Finally, S. C. Johnson's hierarchical clustering analysis[6] was applied to the data quite independently of the multidimensional scaling. This technique groups stimuli (signals) according to their mutual closeness in terms of a distance measure provided by the user. The inter-stimulus distance measure used for these data was defined as follows:

$$d_{jk}^2 = \sum_i (R_{ij} - R_{ik})^2,$$

where d_{jk}^2 is the squared distance between stimuli j and k and R_{ij} is the rating given to the jth stimulus by the ith subject. This measure hopefully reflects the similarity of treatment of two signals by each listener. The computer-implemented Johnson technique produces a hierarchy of clustering of n objects, running all the way from n "clusters" of one object each to one cluster of n objects. It also computes a measure of compactness of the clusters at each level between these extremes. Using this measure, as defined by Johnson,[7] we traced the compactness of various clusters as they grew in size to maximum compactness, in an effort to define types of signals.

2.6 *Results*

The first six dimensions of the MDPREF solution accounted for 28, 22, 6, 4, 4, and 3 percent of the variance, respectively. The large drop after the second dimension suggests that only two dimensions of the solution are interpretable.

Of the parameters used for regression analysis, only modulation frequency (MF) could be located with sufficient confidence to identify it with a dimension of the solution. However, interpretation of the two-dimensional solution as a whole was greatly aided by the cluster analysis. Six clusters, including almost all the stimuli, were found to be at maximum compactness at about the same level in the hierarchy. Inspection of the membership of each cluster suggested a name for the type of signal comprising the cluster. Furthermore, when the stimulus points were projected on the plane of the first two MDPREF dimensions, a fairly simple closed curve could be drawn around each of the clusters without overlapping the others. This projection, along with the cluster contours and the subjects' vectors, is shown in Fig. 1. The pulse-carrier signals are shown as open squares, the sine-carrier AM signals as shaded circles, and the sine-carrier FM signals as filled circles. Subjects' vectors, normalized to unit length in two dimensions*, are shown as arrowheads pointing in the direction of higher evaluation. The vertical axis (Dimension I) corresponds to the MF vector, with low MF (5, 7, 10 Hz) at the bottom of the figure and high MF (20, 40, 80 Hz) toward the top.

The import of the results for calling-signal design is conveyed by consideration of the gross characteristics of the signal clusters. Although a more detailed examination of the results with respect to parameter levels reveals interesting information about tone perception, these findings are deemed inappropriate for present purposes. Complete details are available from the author.[8]

The signals in both clusters below the horizontal axis of Fig. 1 ("gliding pitch" and "trills") are, with three exceptions, FM signals modulated at a slow enough rate that the pitch can be heard to change. In the case of "gliding pitch" signals, the pitch changes in a continuous manner, while the tone of the "trills" jumps from one pitch to another. The three exceptions are AMSC signals of which the single pitch alternates with silence. Note that no listener's vector is located so as to indicate a clear preference for gliding pitch signals, whereas quite a few listeners evaluate trills higher than any other type of signal.

---

* This normalization was adopted in the interest of presenting an uncluttered picture. A similar plot in which overall percentage of variance accounted for was represented by the distance of each arrowhead from the origin revealed no systematic relation between orientation and length of vectors.

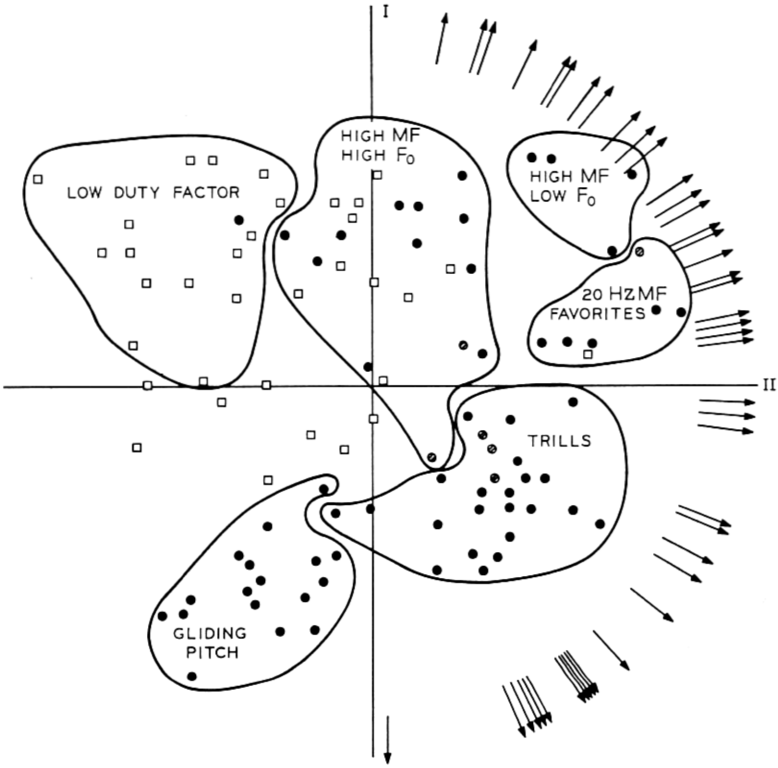

Fig. 1—Projections of 100 signals and 43 listeners' vectors on Dimensions I and II of the MDPREF solution for Experiment 1. Arrowheads represent listeners' vectors, filled circles FMSC signals, shaded circles AMSC signals, and open squares AMPC signals.

The one characteristic shared by all of the signals in the "favorites" cluster is an MF of 20 Hz. They also have low or medium carrier frequencies. This cluster is positioned so as to receive high ratings from many listeners who prefer 10- or 40-Hz MF somewhat more. Its name derives from the fact that it includes the signals with the three highest mean ratings and three others in the top ten. The single AMPC signal to reach this distinction was also first in the earlier study[4]; it appears again in Experiment 4.

The four signals in the "high MF-low F_o" cluster all have MF = 40 or 80 Hz and a carrier frequency of 400 Hz. Some listeners clearly prefer these signals but not many other high-MF signals, to those with lower MF.

The name of the "high MF-high F_0" group is self-explanatory ("high F_0" means carrier frequencies of 1,600 Hz or 800 Hz for AMSC and FMSC signals, and 900 or 700 Hz for AMPC signals). This group receives few high ratings. The last group ("low duty factor") is almost entirely AMPC signals with a low or medium duty factor, which renders harmonics of the modulating frequency prominent. No listener's vector is located so as to indicate a preference for these signals.

2.7 *Conclusions*

Some tentative principles of good calling-signal design suggested by these results are:

(*i*) There seems to be an optimum modulation frequency around 20 Hz, even though individuals vary widely with respect to this parameter. Both the superiority of 20 Hz and the diversity of listeners were observed by Bricker and Flanagan.[4]

(*ii*) Pitch should change abruptly rather than gradually at low modulation frequencies.

(*iii*) Smooth amplitude modulation (high duty factor) is superior to abrupt amplitude modulation (low duty factor).

(*iv*) Low carrier frequency and a narrow, low-centered spectrum are advantages, while high-frequency energy is a disadvantage, whether it arises from a high carrier frequency or the presence of harmonics of the carrier.

The first two principles receive further support from subsequent experiments, the third is not investigated further, and the fourth is explored and refined in the next three experiments.

III. EXPERIMENT 2

3.1 *Background*

The problem most in need of attention after Experiment 1 was the status of pulse-carrier signals. It is clear that most of the AMPC signals and the "gliding pitch" FMSC signals received low ratings, and that the two types are separated in the solution space (Fig. 1). However, they are separated chiefly on Dimension I, which corresponds to modulation frequency, and for very good reason: the AMPC signals in the study all had MF $\geqq$ 10 Hz, while the "gliding pitch" signals all had MF $\leqq$ 10 Hz. The two types project to similar points on Dimension II, which is enigmatic: this dimension could reflect either a perceptual characteristic or merely the confounding in the experimental design. The purpose of Experiment 2, then, was to determine whether AMPC signals required a dimension of their own to describe their perceptual relations with FMSC signals. The physical correlate of the dimension thought to be useful for this purpose is the bandwidth of the signals: AM *sine*-carrier signals have a narrower bandwidth than FMSC signals, which in turn are narrower than AMPC signals. Experiment 2 includes signals of all three types, each represented at the same values of MF so as to remove that source of confounding. Note that a narrow bandwidth was listed as a desirable characteristic in the fourth conclusion of Experiment 1, and that Experiment 2 is designed to provide additional information on this point.

Since this experiment and those that follow use a novel method of collecting data, the procedure and some results it has produced will be briefly reviewed. The technique, called auditory sorting, has been described in the literature.[9] Briefly, it provides the listener with an array of movable pushbuttons, each of which evokes a distinctive sound. The listener arranges the buttons (sounds) in groups or in order according to instructions. In an early experiment with this technique, listeners were asked to group 24 three-parameter frequency-modulated tones according to similarity. Using an appropriate multidimensional scaling technique, it was possible to recover a perceptual space that closely resembled that recovered from the much more laborious pair-comparison procedure in a companion experiment.[10] In another experiment, listeners ordered a subset of the tones used in Experiment 1 according to preference. From these data, MDPREF constructed a space very much like that based on the rating data for the same subset. The strategy in this experiment, then, was to include enough FMSC signals to recover a three-dimensional perceptual space and then observe whether the AMPC signals, the AMSC signals, or both required an additional dimension to account for their evaluations.

3.2 *Signals*

Twenty-four signals, all of which were derived from an 800-Hz carrier frequency, were used in this experiment. They were of three types: FMSC ($n = 18$), AMSC ($n = 3$), and AMPC ($n = 3$). Each type was represented at the same three modulation frequencies: 10 Hz, 20 Hz, and 40 Hz. There were six FMSC signals at each MF, realizing all combinations of two FM waveforms (sinusoidal and rectangular) and three amounts of frequency deviation (± 3, ± 10, and ± 25 percent). The AM signals were modulated with a symmetrical sinusoidal envelope, and the PC signals employed a carrier with approximately equal-amplitude components at 800, 1,600, and 2,400 Hz.

3.3 *Listeners*

Thirty Bell Laboratories employees, 17 male and 13 female, served as listeners in this experiment.

3.4 *Procedure*

The listener was seated in a small sound-attenuating booth with the sorting apparatus in front of him and a loudspeaker on the wall above it. He was shown how each button evoked a different sound (signal) from the speaker, and instructed to arrange the signals (buttons) from right to left according to "how much [he] would like each of these tones if it replaced the telephone bell." Listeners were permitted to produce partial orderings, i.e., to arrange the signals in groups such that the evaluative ordering obtained between groups, but signals within a group were tied.

3.5 *Analysis*

The vectors were each normalized and arranged in a 30 (listeners) by 24 (signals) matrix to serve as input for MDPREF. Regression techniques were used to find, in the resulting space, three orthogonal directions that best corresponded to the three parameters of the FMSC signals. A fourth dimension was located so as to satisfy three conditions: (*i*) mutual orthogonality to the first three, (*ii*) maximization of residual variance accounted for, and (*iii*) close nonlinear correspondence to a property defined by three levels of spectral width—narrow (AMSC), medium (FM), and wide (AMPC).

3.6 *Results*

In the figures that present the results, plotting symbols have been used that suggest the parameter values of the signals they represent. Thus, large symbols are used for 25-percent frequency modulation, medium for 10-percent, and small for 3-percent; round symbols for sinusoidal FM, square for rectangular FM. Modulation frequency in Hz is given by arabic numerals. For the AM signals, sinusoidal and pulse carriers are distinguished by symbols representing one cycle of their respective waveforms.

The projections of all 24 signals on the plane of the first two solution dimensions are shown in Fig. 2a, and their projections on the plane of the third and fourth dimensions are shown in Fig. 2b. For the 18 FMSC signals, the first three dimensions represent modulation frequency, modulation percentage (MP), and modulation waveform (WF), re-

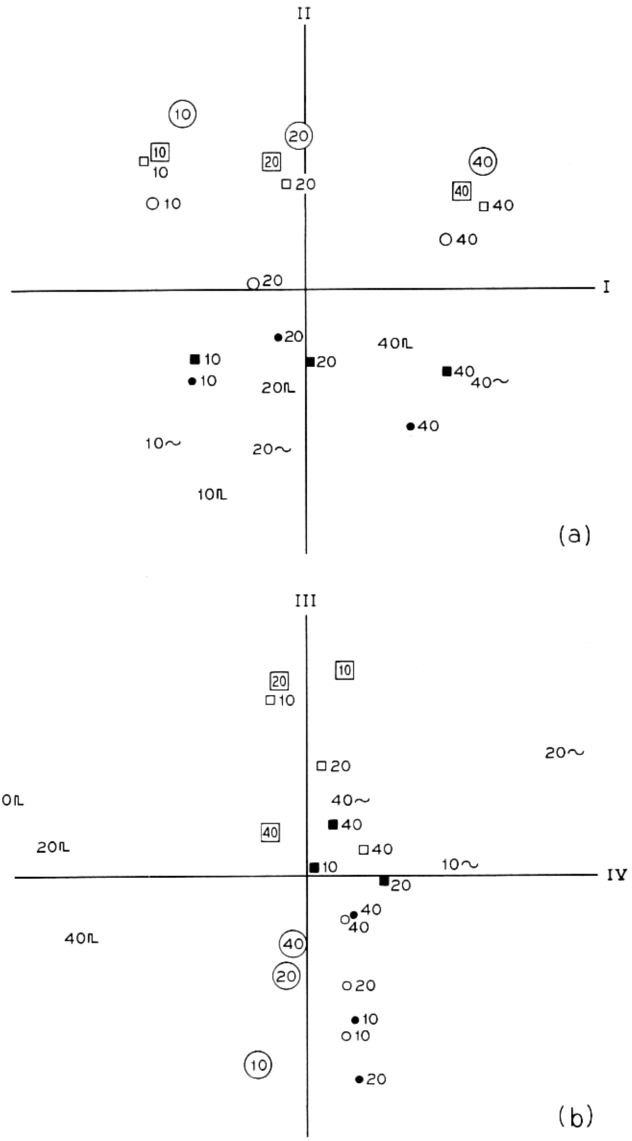

Fig. 2—Projections of 24 signals on the MDPREF solution dimensions for Experiment 2: (a) Dimensions I and II; (b) Dimensions III and IV.

spectively. This configuration resembles that obtained from the aforementioned similarity-judgment experiments[10] in sufficient detail to support the conclusion that the present experiment reveals dimensions of perceptual significance.

Note that the AM signals fall in appropriate places on Dimension I (MF) and that they are just beyond the 3-percent FM signals on Dimension II (MP); this latter location is consistent with their having 0 percent frequency deviation. The chief function of Dimension IV is to separate the AM signals, with the AMPC signals to the left and the AMSC signals to the right. This dimension accounts for 10 percent of the variance, which compares favorably with 28, 13, and 16 percent for the first three, respectively.

The listeners' vectors are shown in Fig. 3a and b. Whereas Fig. 3a shows the usual diversity of opinion with respect to MF (and MP as well), Fig. 3b shows a considerable concentration of vectors so as to reflect low ratings for both AMPC and low-rate sinusoidal FM signals. It is not clear from these figures whether evaluation continues to improve as bandwidth is reduced. Subjects are in fact evenly divided as to whether their highest-ranked AMSC signal is ranked above their highest-ranked FMSC signal, and mean normalized rank is slightly higher for FM. Thus, although bandwidth is established as a perceptually significant parameter, listener evaluation is not a monotonic function of it.

3.7 *Conclusion*

A reasonable accounting of the data demands that AM signals be regarded as differing from the (three-dimensional) FM signals on a fourth dimension. Although signal bandwidth provides a satisfactory interpretation of this dimension, its relation to listener evaluation is other than monotonic.

IV. EXPERIMENT 3

4.1 *Background*

While the experiments to this point had considerably clarified the nature of the perceptual space, they had not explored all combinations of parameter values. In particular, Experiment 2 suffers from confounding of carrier type with modulation type (FM or AM), even though it served to un-confound these parameters with modulation frequency. Thus we are left with the formal possibility that Dimension IV of Experiment 2 could be interpreted as "modulation type" rather

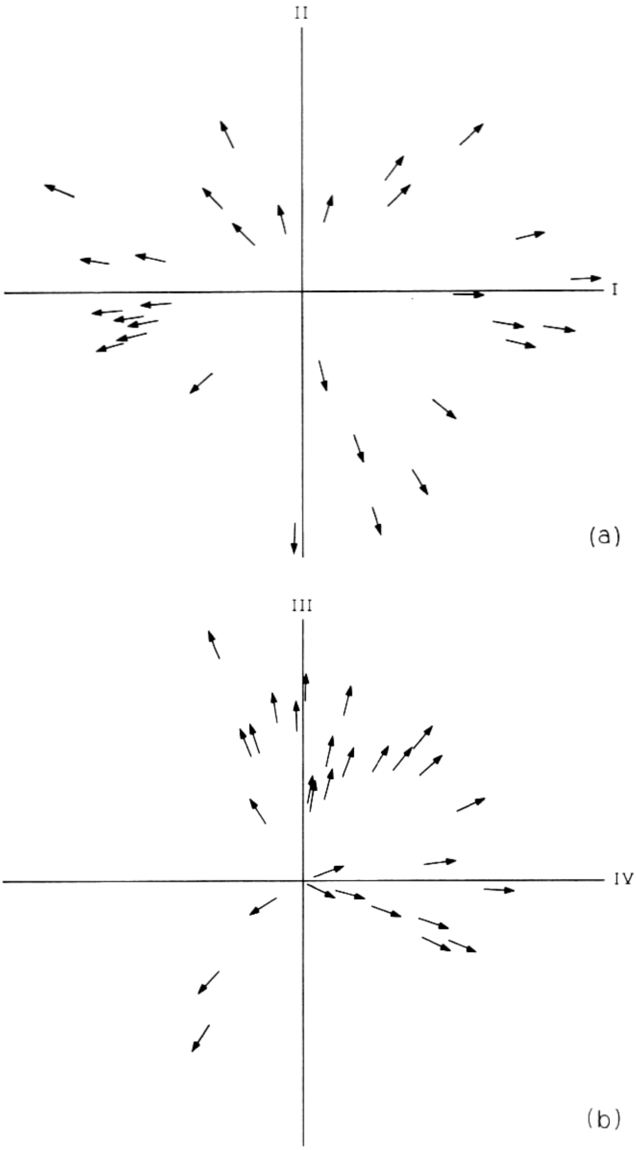

Fig. 3—Locations of listeners' vectors with respect to the dimensions shown in Fig. 2: (a) Dimensions I and II; (b) Dimensions III and IV.

than "bandwidth," and with the practical possibility that pulse-carrier signals might receive higher ratings if frequency modulated. Experiment 3 allows for complete factorial arrangement of both types of carrier with both types of modulation.

4.2 *Signals*

All possible combinations of three modulation frequencies (10, 20, and 40 Hz), two modulation types (AM and FM), and two carrier types (sinusoidal and pulse) were generated, for a total of twelve signals. The carrier frequency was 800 Hz, and the FM modulating waveform was rectangular with ±10-percent deviation. The pulse carrier contained the first three harmonics of the carrier and the AM parameters were as in Experiment 2.

4.3 *Listeners*

Forty-four Bell Laboratories employees served as listeners. They were each selected with approximately equal frequency from two categories each of age (over or under 30) and sex. In addition, each listener submitted to an audiometric screening test; none was found to exhibit a clinically significant loss in the range of the signals under test.

4.4 *Procedure*

Each listener produced an evaluative ordering or partial ordering of the 12 signals, using the sorting apparatus under the same instructions as in Experiment 2. At the termination of the ranking procedure, each listener was asked to state whether he liked "the bell on his home telephone" more than none, some, or all of the tones, and to locate it in the hierarchy if "some".

4.5 *Analysis*

The objective in this and subsequent experiments was more to identify optimum parameter combinations than to discover perceptual dimensions. Consequently, the results are presented in terms of conventional statistical summaries and tests of significance rather than multidimensional analysis. The basic datum is the rank assigned each of the 12 signals by each of the 44 listeners.

4.6 *Results*

The median rank assigned each of the 12 signals is shown in Table I, where lower numbers indicate higher evaluations. It is immediately apparent that sine carrier ranks higher than pulse carrier in each of the six comparisons with the other factors held constant. Neither of the

TABLE I—MEDIAN RANK ASSIGNED EACH OF TWELVE SIGNALS, EXPERIMENT 3

Modulation Type	FM		AM	
Carrier Type	Sine	Pulse	Sine	Pulse
Modulation Frequency (Hz) 10	4.5	5.6	5.5	7.5
20	2.8	6.9	3.8	6.9
40	5.8	9.4	4.7	9.3

other parameters shows quite so consistent an effect. An analysis of variance showed all three parameters to have statistically significant effects, with carrier type the largest and modulation type the smallest. Again 20 Hz is the highest-ranking modulation frequency; the rank distributions for most of the 10-Hz and 40-Hz signals reflect the diversity of opinion about MF observed in the earlier experiments, in that they are generally bimodal or broadly dispersed. Overall, FM is somewhat higher ranked than AM. Another finding of the analysis of variance was that there was no systematic difference among the four age-sex groups in their patterns of evaluation.

Table II shows, for each signal, the number of listeners who ranked it above or equal to their remembered concept of the bell. The pattern of preferences here is much the same as that shown by median rank (in fact, it would be statistically the same if there were no correlation between electronic signal evaluation and "bell" evaluation). The main value of this measure is to give the signal ratings some external reference, however tenuous.

4.7 *Conclusion*

The data show clearly that at least for a carrier frequency of 800 Hz, a signal containing three harmonics of the carrier is less well liked than its single-frequency counterpart, regardless of whether amplitude or frequency modulation is employed. Furthermore, there is no practi-

TABLE II—NUMBER OF SUBJECTS (OUT OF 44) RANKING EACH SIGNAL ABOVE OR EQUAL TO THE "BELL", EXPERIMENT 3

Modulation Type	FM		AM	
Carrier Type	Sine	Pulse	Sine	Pulse
Modulation Frequency (Hz) 10	11	12	17	7
20	23	11	19	5
40	14	5	12	4

cal advantage to frequency-modulating a pulse-carrier signal, although FMSC signals are again slightly superior to AMSC signals.

Discovery of a way to improve evaluations of pulse-carrier signals would be valuable, because such broadband signals have a well-established advantage in detectability. A search for such a means gave rise to the next experiment which although brief and unavailing was informative in other respects.

V. EXPERIMENT 4

5.1 *Background*

The best-liked FM signal (20-Hz, 10-percent rectangular FM) has a musical aspect worth noting: the two alternating frequencies (720 Hz and 880 Hz) stand in a relation close to a major third, generally thought to be a pleasing musical interval. The pulse-carrier signals used so far also have a musical aspect: the three components (e.g., 800, 1,600, and 2,400 Hz) establish two intervals—an octave and a major fifth. While not dissonant, the fifth is generally considered harsher and less pleasing than the third. But certain members of the harmonic series other than the first three can be chosen so as to generate thirds (fourth and fifth harmonic) and pleasing inverted (or open) triads. This experiment was an attempt to improve the rating of pulse-carrier signals by selecting such combinations.

5.2 *Signals*

Each of the eight signals in this experiment had three components whose relative amplitudes decreased at a rate of 3 dB per octave from 500 to 4,000 Hz. All signals were amplitude modulated as before at 20 Hz. The frequencies of all the components in kHz are shown in Table III. The more important musical aspects of this set are: (*i*) Signals 1, 2, 3, and 6 are *compact,* in that adjacent harmonics of the respective fundamentals are selected. Of the compact signals, only number 3 represents a major triad (second inversion); (*ii*) Signals 4, 5, 7, and 8 are *open,* in that certain harmonics are suppressed in between those that are passed. Each of the open tones constitutes a major triad, inverted and opened to span more than an octave.

5.3 *Listeners*

Bell Laboratories employees were selected to represent wide variation in musical skill and training, and to have normal hearing over the range of component frequencies. The experiment was terminated after six listeners had been run.

TABLE III—FREQUENCIES IN KHz OF COMPONENTS OF EIGHT SIGNALS, EXPERIMENT 4

Signal No.	1	2	3	4	5	6	7	8
First Component	0.5	1.0	1.5	1.5	0.5	0.75	1.0	1.25
Second Component	1.0	1.5	2.0	2.5	1.5	1.5	1.25	2.0
Third Component	1.5	2.0	2.5	4.0	2.5	2.25	3.0	3.0

5.4 *Procedure*

Listeners were asked to use the sorting apparatus to arrange the eight signals in evaluative order, as before.

5.5 *Results*

Regardless of the musical background, no subject ranked any of the five triad-producing signals first. The signal most often ranked first was number 1, which had the lowest frequencies for all three components. Number 6 was the only other signal ranked above the median by all listeners. Listeners described the low-ranked "musical" signals as "high-pitched," "tinny," and "jarring." One listener who was sophisticated in both music and acoustics recognized the differences in musicalness, but averred that the high-frequency components were irritating, even though they served to complete a chord. Since even those listeners expected to be most favorably disposed toward chord-signals rejected them in favor of signals with low component frequencies, the experiment was terminated after only six listeners.

In addition to rejecting the notion that musicalness of component intervals might "rescue" broad-spectrum signals, this experiment affords certain informative comparisons with earlier studies. For example, the first three signals have the same bandwidth, but differ in location of their spectra. Rankings plummet from first through average to last as frequency of components increases across this set. Signal number 6 has a greater bandwidth than any of these and a highest component between those of signals 2 and 3, yet its overall rank was equivalent to that of signal number 1. These facts suggest that listener evaluation depends in a complex way on the frequencies of the components, so that both the average frequency and the highest frequency can be determining. Bandwidth per se, it seems, is much less important. Searching through the details of Experiment 1 produces support for this notion, and further suggests that the region between 2,000 and 2,500 Hz is critical for upper component, 1,500 to 2,000 Hz for average. One would interpret the poor showing of pulse-carrier signals in Experiments 2 and 3 in retrospect as an invasion of a critical frequency region by the highest component, rather than as an effect of bandwidth or the mere presence of harmonics.

It happens that signal number 2 in Experiment 4 has exactly the same specifications as the lone AMPC signal to join the "favorites" cluster in Experiment 1 (Fig. 1). Since signal number 2 was outranked in Experiment 4 by signal number 6, even though the latter has a component above 2,000 Hz, we might expect signal number 6 to compete well with sine-carrier signals. However, the most similar signal in Experiment 3 (AMPC, 20 Hz, components at 800, 1,600 and 2,400 Hz) was considered equal to or better than "the bell" by only 5 out of the 44 listeners, compared to 23 out of 44 for the best FMSC signal (see Table II). Although it is possible that three harmonics of 750 Hz (especially with 3 dB per octave attenuation) could be much more pleasant than three harmonics of 800 Hz, it is safer, in the absence of a complete map of component-frequency effects, to take these Experiment 3 findings as a guide to how signal number 6 would fare against FMSC signals. The reason for the attention given here to signal number 6 is a practical one: its acoustic specifications are exactly the same as those of a ringer now under development.[11] This ringer has been shown[12] to be satisfactorily detectable in typical room noise, and to be superior in this respect to a narrow-spectrum FMSC signal. The present experiments suggest that while a three-harmonic 750-Hz signal is a good one of its type, it is likely to be less well liked by listeners than a good FMSC signal.

Since the laboratory affords no way to equate pleasantness and detectability, evaluation of both leading signals under operating conditions seemed an appropriate means of resolving the conflict between the criteria. To make this evaluation possible, the Telephone Laboratory at Indianapolis modified its basic design so that it could generate an FM signal. Questions that arose in the course of this redesigning effort prompted the next and last experiment.

VI. EXPERIMENT 5

6.1 *Background*

The ringer consists mainly of an electromagnetic transducer and a resonant cavity. The resonator must be small enough to fit inside a telephone station set and large enough to resonate the lowest frequency component of the desired signal. The higher the carrier frequency, the smaller the ringer could be, so the listener-evaluation function of carrier frequency (F_o, or average) is important design information Experiment 1 had indicated that signals with F_o = 1,600 Hz were not as well liked as those with F_o = 800 Hz, but there was also a suggestion that evaluation was not monotonic with F_o. In any event, the parameter F_o had not been explored in sufficiently small steps to guide the design of a ringer for narrow-band signals.

Another purpose of Experiment 5 was to assess the effects of superimposed amplitude modulation on listener evaluation. This question arose because such amplitude modulation was found to be technically difficult to eliminate from the design under consideration.

6.2 *Signals*

There were eight FMSC signals involved in this experiment. Each was rectangularly frequency-modulated at 20 Hz, with the upper and lower components standing in the ratio of 5 to 4 in frequency—a major third. There were four values of F_o, as shown in Table IV, along with the upper and lower component frequencies. There were two signals at each F_o, one of which was pure FM, the other of which was amplitude modulated at 20 Hz in such a way as to imitate the effect found in the practical design.

6.3 *Listeners*

Thirty listeners, 10 male and 20 female, were recruited from among Bell Laboratories' clerical, shop, and technical employees.

6.4 *Procedure*

Each listener used the auditory sorting apparatus to rank or partially rank the eight signals, as in Experiments 3 and 4.

6.5 *Results*

The number of listeners who assigned each rank to each of the eight signals is shown in Fig. 4. This method of presenting the data is resorted

TABLE IV—FREQUENCIES IN Hz OF F_o AND BOTH COMPONENTS OF SIGNALS TESTED IN EXPERIMENT 5

F_o	Upper Component	Lower Component
1,350	1,500	1,200
1,125	1,250	1,000
900	1,000	800
675	750	600

to here because the extreme biomodality of the ranking for $F_0 = 675$ Hz renders any measure of central tendency misleading. For practical purposes, one would wish to avoid a signal about which listener opinion is so divided. There is more of a consensus on low ranks for the signals with $F_o = 1{,}350$ Hz, and there is little to choose among the signals with $F_o = 1{,}125$ or 900 Hz.

Detailed analyses confirm the impression given by the figure that there is little difference between signals with or without superimposed AM. A tally was made of the outcome of each of the four same-carrier-frequency comparisons for each listener, the possible outcomes being FM > AM, AM > FM, and FM = AM. The results were that AM made essentially no difference to 11 listeners, while 6 listeners preferred FM only three or four out of four times, and 5 listeners preferred AM to FM only three or four times.

The results of this experiment are viewed as supporting two engineering decisions taken subsequently, rather than as evidence for more general conclusions. The first was to develop a narrow-band FM ringer with $F_o = 1{,}012.5$ Hz, which is midway between the two generally acceptable F_o in the experiment. The second was not to attempt to eliminate the superimposed AM, in the light of the indifference to it apparent in the experiment.

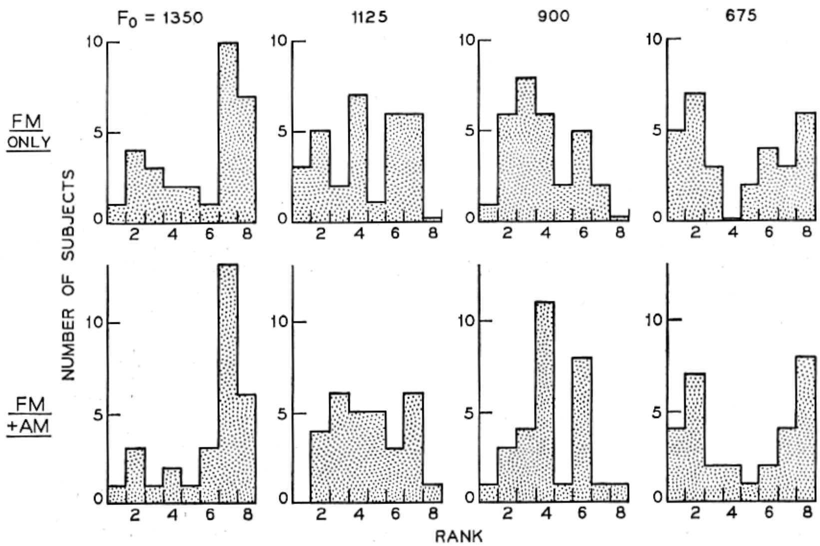

Fig. 4—Histograms of number of listeners assigning each of eight ranks to each of eight signals, Experiment 5; $N = 30$ for each histogram.

VII. SUMMARY OF CONCLUSIONS

Results of these experiments permit the specification of two ringers—one narrow-band and one broad-band—that are promising with regard to listener evaluation. The narrow-band signal is frequency modulated at 20 Hz, with its upper and lower components tuned to a frequency ratio of 5 to 4, averaging around 1 kHz. The broad-band signal is amplitude modulated at 20 Hz, and has three roughly equal-amplitude components at 750, 1,500, and 2,250 Hz. The evidence indicates that the narrow-band signal will be preferred to the broad-band signal by a majority of listeners when the two are in direct contest, and that the broad-band signal is more detectable in typical room noise when the two are equated for power.

Laboratory studies do not reveal, however, how listeners will evaluate either signal after some experience with it in actual use. Neither do they tell us how effective these signals will be in practice. A field trial involving residential subscribers is planned to gather information on these points. This trial also makes it possible for subscribers to tell us something about the tradeoff between the pleasantness advantage of the narrow-band signal and the detectability advantage of the broad-band signal by adjusting their volume controls. Informal experiments have indicated that listeners attenuate an unpleasant signal, when given a volume control in the laboratory. If subscribers behave similarly in the field, we shall have as one of the data the amount by which they offset the detectability advantage of the less pleasant signal. In any event, we shall collect data on answering times, no-answer rates, and opinions-after-experience. As useful as the laboratory studies have been, they could not have provided information on these crucial points.

The field trial will also provide an opportunity for a direct comparison of these two tone ringers with a widely used gong (C4). There are so many differences between tone ringers and gong ringers, ranging from the obvious difference in excitation to the factor of familiarity, that interpretation of this aspect of the study will be difficult. Nevertheless, these results may require modification of some of the principles (e.g., those pertaining to bandwidth) derived from this series of experiments on tone ringers alone.

VIII. ACKNOWLEDGMENTS

For their contributions to one or more phases of this work, the author is grateful to J. D. Carroll, Jih-Jie Chang, J. L. Flanagan, Barbara McDermott, Sandra Pruzansky, and M. Wish.

REFERENCES

1. Archbold, R. B., Ithell, A. H., and Johnson, E. G. T., "The Ideal Characteristics for the Calling Signal of a Subscriber's Telephone Set," Research Report No. 21143, Post Office Research Station, Dollis Hill, London, November 22, 1967.
2. Mevissen, H. M. J., and Stremmelaar, H., "A Study on the Appreciation of Tone Ringing by the Subscribers of a Fully Electronic Exchange," Fourth Int. Symp. on Human Factors in Telephony, Bad Wiessee, Germany, September 22–27, 1968.
3. Gale, J., "Human Factors and the Telephone," Northern Electric Telesis, *1*, No. 9 (October 1970), pp. 287–293.
4. Bricker, P. D., and Flanagan, J. L., "Subjective Assessment of Computer-Simulated Telephone Calling Signals," IEEE Trans. Audio and Electroacoustics, *AU-18*, No. 1, (March 1970), pp. 19–25.
5. Chang, J. J. V., and Carroll, J. D., "How to Use MDPREF, a Computer Program for Multidimensional Analysis of Preference Data," unpublished work.
6. Johnson, S. C., "Hierarchical Clustering Schemes," Psychometrika, *32* (1967), pp. 241–254.
7. Johnson, S. C., "A Simple Cluster Statistic," unpublished work.
8. Bricker, P. D., Carroll, J. D., McDermott, B., Pruzansky, S., and Wish, M., unpublished work.
9. Bricker, P. D., Johnson, S. C., and Mattke, C. F., "Apparatus for Auditory Stimulus Sorting," Behaviorial Research Methods and Instrumentation, *1* (1969), pp. 148–149.
10. Bricker, P. D., and Pruzansky, Sandra, "Comparison of Sorting and Pairwise Similarity Judgment Techniques for Scaling Auditory Stimuli," J. Acoust. Soc. Amer. *47* (1970), p. 96 (A).
11. Hunt, R. M., "Determination of an Effective Tone Ringer Signal," Preprint No. 722, 38th Conv. Audio Eng. Soc., Los Angeles, May 1970.
12. Cooper, P. T., unpublished work.