Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters*

M. R. Schroeder, D. Gottlob, and K. F. Siebrasse

Drittes Physikalisches Institut, Universität Göttingen, 34 Göttingen, Federal Republic of Germany
(Received 1 October 1973; revised 5 March 1974)

We describe a subjective preference evaluation of European concert halls and attempt to correlate the subjective data with objective (geometric and acoustic) parameters of the halls. For the subjective evaluations, reverberation-free magnetic-tape recordings of a symphony orchestra are played from the stage of the hall and recorded at the "eardrums" of a specially designed dummy head with realistic surface impedance of the ear canal and pinnae. These recordings are electronically processed such that, when played back over two selected loudspeakers in an anechoic chamber, the recorded signals are recreated at the eardrums of a human listener. This method of recording and reproduction makes possible instantaneous comparison of the acoustic qualities of different halls under realistic free-field listening conditions on the basis of identical musical source material. Listeners of differing degrees of musical training make preference judgments of musical quality in paired-comparison tests. The raw preference data are subjected to multidimensional factor analysis yielding one "consensus preference" factor and several "individual difference" factors. Results are reported of correlations of the consensus preference with *objective* parameters of the halls (volume, width, reverberation time, initial time-delay gap, "definition," and "interaural coherence").

Subject Classification: 55.45, 55.20.

INTRODUCTION

Judgments of acoustical quality are difficult and subtle. Meaningful comparisons between different concert halls are arduous because of a listener's inability to transpose himself from one hall to another within his relatively short acoustic-memory span. In addition, there is the problem of trying to form one's judgments on comparisons of different musical performances. The fact that the *program* may list the same compositions is of little help when trying to judge details in the perceived sound that are perhaps more influenced by the manner of the musicians' playing their instruments or by the orchestra's seating arrangement than by the objective characteristics of the hall in which the music is being played.

An obvious way out of this dilemma is to record the program material in different concert halls by means of microphones at the "eardrums" of a *dummy head*[1,2] and then to play these recordings back at a convenient time and place, allowing for rapid switching between different halls. To ensure identity of musical program material, a *recording* of an orchestra should be used and played from the stage—rather than repeated musical performances. Naturally, the orchestra should be recorded in an acoustically neutral environment; in other words, in an anechoic chamber. Fortunately, such recordings are available, even from full-scale symphony orchestras.[3]

The next question is how to present the dummy-head recordings to the human listener for evaluation. Putting the two dummy-head signals on high-quality stereo earphones worn by the human listener is one perfectly legitimate method of reproducing the recorded sound. But this method introduces other problems. It is well known that listening over earphones does not properly recreate the perceived acoustic "space" or the sense of being surrounded by sound, which is one of the important acoustic qualities that we wish to describe in an objective way.

I. THE RECREATION OF A SOUNDFIELD AT THE TWO EARS

For the reproduction of the recorded program material, we used a method originally proposed and demonstrated by Schroeder and Atal.[4] The idea is to record with a dummy head and mix and prefilter the dummy-head signals in such a way that, when they are radiated from two loudspeakers in front of the listener, the signal from the dummy's right ear will *only* go to the listener's right ear and that from the dummy's left ear will go *only* to the listener's left ear, just as in earphone listening but with the proper free-field coupling of the ear canal and the desired invariance of the perceived acoustical space when the listener's head is rotated around a vertical axis.

For simplicity, one may think about this idea in the following terms: The sound radiated from each loudspeaker goes into *both* ears and not just the "near" ear of the listener, as would be desired. In other words, there is "cross talk" from each loudspeaker to the "far" ear. However, by radiating properly mixed and filtered compensation signals from the loudspeakers, the unwanted "cross-talk" signals can be cancelled out.

Figure 1 illustrates the result of sound diffraction and "cross talk" around the human head. The function $s(t)$ is the impulse response from the terminals of one loudspeaker in an anechoic chamber to the "near" ear of a dummy sitting at some distance from the loudspeaker; $a(t)$ is the impulse to the "far" ear.

One notices both a *delay* and an *attenuation* in the far-side signal $a(t)$ relative to the near-side signal $s(t)$—a result of the diffraction and shadowing effect of the head.

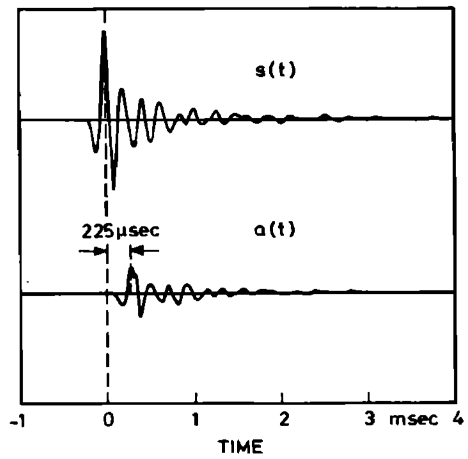

FIG 1. Impulse responses between loudspeaker terminals and dummy head microphones: $s(t)$ = near-side response; $a(t)$ = far-side response. Loudspeaker in front of dummy at 36° to one side and a distance of 2.5 m in anechoic chamber.

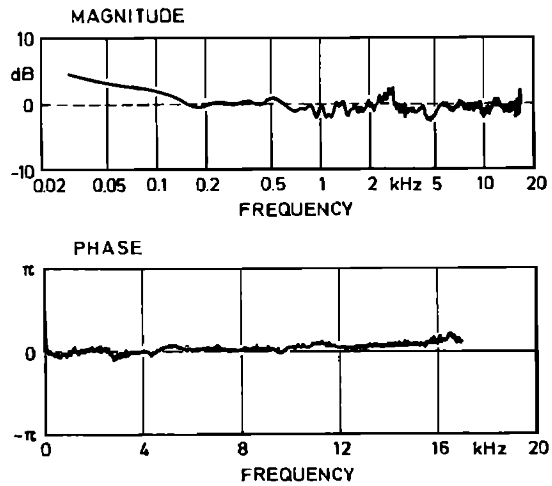

FIG. 3. Amplitude ratio and phase difference between the two loudspeakers used in the experiment. The phase difference is plotted on a linear frequency scale.

Figure 2 shows the Schroeder–Atal filtering scheme to cancel the cross talk around the head. Here, $S = S(\omega)$ is the Fourier transform of $s(t)$ and $A = A(\omega)$ that of $a(t)$. $C = C(\omega)$ is an abbreviation for $-A(\omega)/S(\omega)$. The circles labeled $1/(1-C^2)$ and $1/S$ represent filters with the indicated transfer functions ("amplitude and phase responses").

For a signal applied to the right input ("R" at the upper right), the right-ear signal becomes, in terms of its spectrum:

$$Y_R = R \cdot [(1-C^2)^{-1} \cdot S^{-1} \cdot S + C(1-C^2)^{-1} \cdot S^{-1} \cdot A] , \tag{1}$$

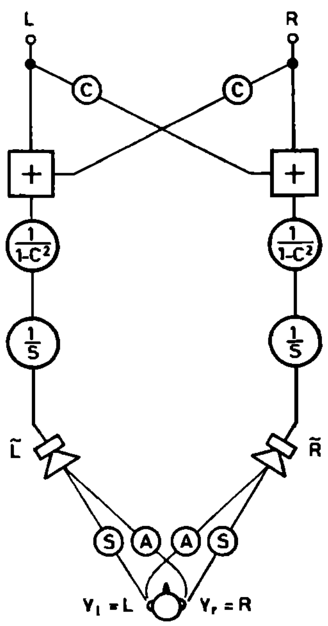

FIG. 2. Prefiltering scheme for free-field reproduction method. $S = S(\omega)$ and $A = A(\omega)$ are nearside and farside transfer functions, respectively, between loudspeakers and eardrums. The circles labelled $1/S$, $1/(1-C^2)$, and C are filters with corresponding frequency responses. Abbreviation: $C = C(\omega) = -A(\omega)/S(\omega)$.

or, with $C = -A/S$,

$$Y_R = R ,$$

as desired.

The left-ear signal becomes

$$Y_L = R \cdot [(1-C^2)^{-1} \cdot S^{-1} \cdot A + C \cdot (1-C^2)^{-1} \cdot S^{-1} \cdot S] = 0 , \tag{2}$$

also as desired.

Because of the assumed symmetry, this cross-talk cancellation occurs likewise for a signal applied to the left input ("L", upper left), and because of linearity, for signals applied to both inputs. Thus, two arbitrary signals applied to inputs R and L will reappear at the listener's eardrums. Since the listening condition is a free-field one, proper localization is guaranteed. In fact, even elevation angles are properly perceived, including sound-source locations in the *median* plane. Ideally, the listener's head should be fixed. Yet, we found that, for small head movements (about ±10°), the externalized sound image remains stationary in a laboratory reference frame (i.e., it does not turn with the head as in earphone listening).

For this scheme to work properly, the loudspeakers must have nearly identical amplitude and phase responses. Figure 3 shows the amplitude and phase *difference* of the frequency responses of the two loudspeakers we selected for our listening tests. These differences are negligibly small (except possibly below 100 Hz).

Figure 4 shows $S(\omega)$, the near-side amplitude and phase responses. The peaks and valleys result predominantly from sound diffraction at the human head (and, to a lesser extent, from the loudspeaker response).

Figure 5 shows the inverse of $S(\omega)$ as an example of one of the filter responses needed in the sound-reproduction system. For the inverse to be realizable (either on the computer or as an analog filter), it must

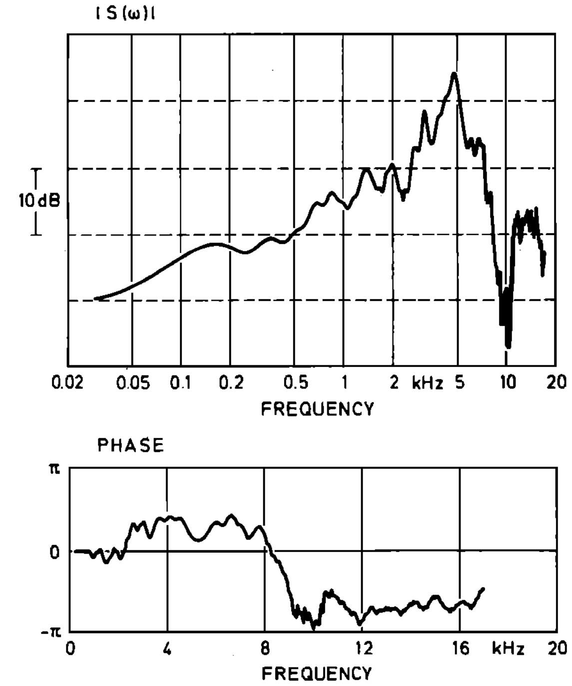

FIG. 4. Fourier transform $S(\omega)$ of $s(t)$ from Fig. 1, i.e., the transfer function between one of the loudspeakers and the nearside eardrum. Upper panel: amplitude response in decibels vs logarithmic frequency. Lower panel: phase in radians as a function of *linear* frequency.

first be bandlimited. In our case, the bandlimiting extends from about 100 Hz to 15 kHz.

In order to test the validity of the prefiltering method, we went through the following sequence of steps:

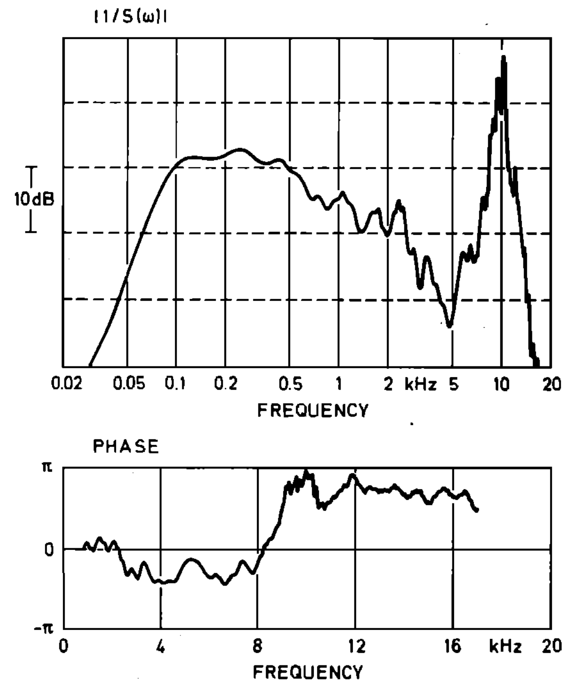

FIG. 5. Bandlimited inverse response of $S(\omega)$ from Fig. 4. This is an example of a response employed in a prefiltering scheme (circles labelled $1/S$ in Fig. 2).

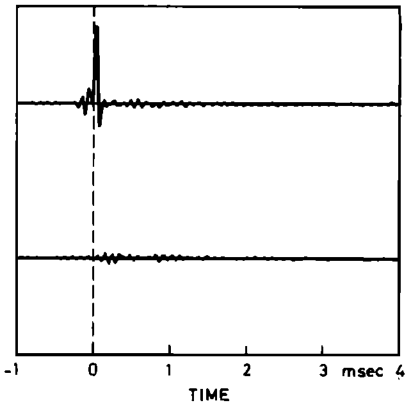

FIG. 6. Result of test of prefiltering scheme. Upper panel: nearside response; lower panel: farside response. The farside response is on the average 25 dB below the nearside response between 200 Hz and 7 kHz. Nearside response has a flat amplitude response (within ± 1.5 dB) between 200 Hz and 7 kHz and is phaselinear within $\pm 20°$ between 70 Hz and 7 kHz.

(1) We applied a short electrical impulse to one of two loudspeakers located in an anechoic chamber and recorded the resulting microphone signals from the ears of a dummy head at some distance in front of the loudspeakers.

(2) From the two recorded impulse responses, we computed the appropriate filter responses for the compensation scheme (Fig. 2) and approximated these filter responses by digital filters.

(3) We then filtered an electrical impulse of about 25-μsec duration by these filters to produce a pair of "loudspeaker signals."

(4) We then placed the dummy *again* in front of the loudspeakers in the anechoic chamber with a *misalignment* typical of that of a *human* listener in our subjective evaluation tests and we radiated the "loudspeaker signals" and recorded from the dummy's microphones.

For perfect reproduction, we should now have recorded, from the dummy's right ear, a sharp impulse and *no* signal at all from its left ear.

Figure 6 shows the results: the upper trace is the right-ear signal as a function of time, the lower trace the left-ear signal. The result looks reassuring (i.e., the left-ear signal is appreciably smaller than the right-ear signal).

In order to gain a better *quantitative* understanding of these two signals, we Fourier transformed them on the computer with the following results:

(1) The right-ear signal has a spectral amplitude which is flat within ± 1 dB between 200 Hz and 7 kHz;

(2) the phase is linear within $\pm 10°$ between 100 Hz and 3 kHz; and, most importantly,

(3) the left-ear signal averages 25 dB below the right-ear signal between 200 Hz and 7 kHz.

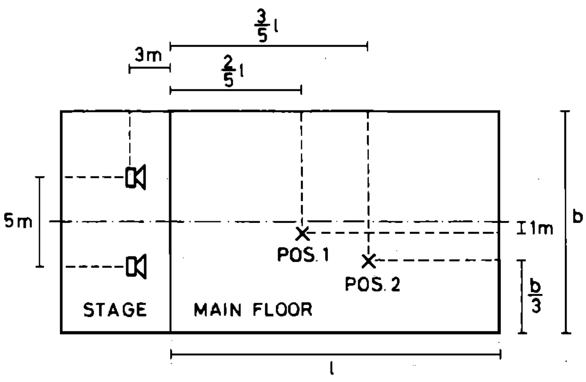

FIG. 7. Loudspeaker and dummy-head recording positions in rectangular hall (similar configurations were used in halls of other shapes).

II. THE CONCERT-HALL MEASUREMENTS AND THE LISTENING TESTS

Figure 7 shows the typical loudspeaker and recording positions in a rectangular hall. (Similar arrangements were used in fan-shaped halls.)

The test tape is a two-track recording of the 4th movement of Mozart's Jupiter Symphony made by the English Chamber Orchestra. We radiated selected portions of this tape from the stage, as shown, to simulate (in a crude way) the spatial extent of an orchestra.

Figure 8 illustrates the listening- and subjective-evaluation phase of our experiment. Two concert halls are compared at a time, the listener being allowed to switch back and forth between the two recordings as often as he wishes and even to ask for replays.

The listener has to indicate which of the two halls he prefers, if either. The score for a preferred hall is 1; for the other it is -1. For no preference, both halls are given a score of 0. Thus, ours is a simple *preference* test—distinct from *similarity* judgments[5] which likewise reveal underlying subjective factors but without the preference information considered crucial in our task.

The preference scores for all subjects and halls are accumulated in a preference matrix whose entries indicate how many times each hall was preferred by each listener.

This matrix was then subjected to a metric linear factor analysis.[6] For our preference data, the factor analysis yielded two or, at most, four significant factors. The variances accounted for by the first four factors depend somewhat on which halls are included in the test. In most cases, the relative variances accounted for by the first four dimensions were found to be about 50%, 15%, 10%, and below 10%, respectively. Thus, we have one factor of overriding significance (50%), another moderately significant factor (15%), and two additional factors of doubtful significance ($\leq 10\%$). The remaining factors are insignificant by several standard tests.

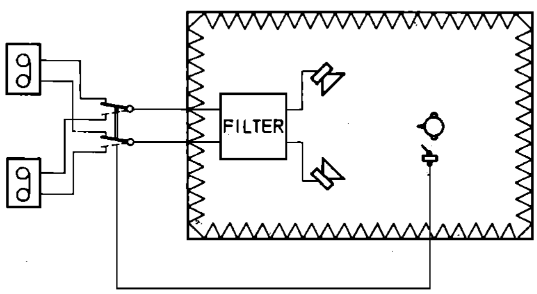

FIG. 8. The setups for the subjective comparisons of the concert halls.

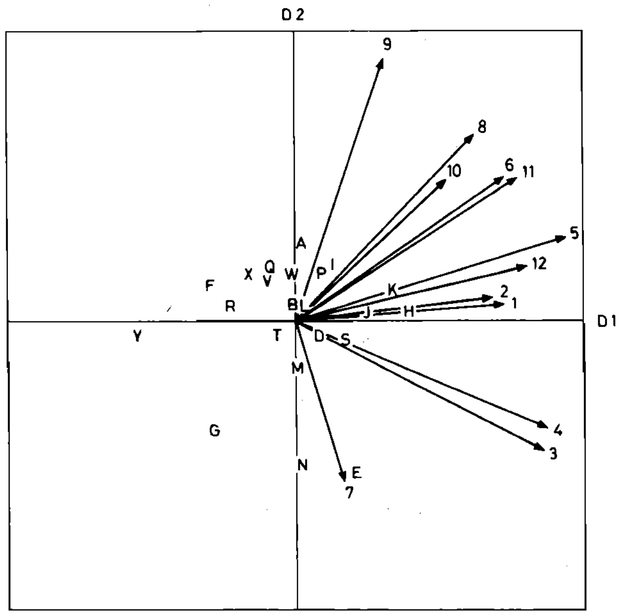

FIG. 9. Preference space: the result of a factor analysis of individual preference judgments. Dimensions 1 and 2 (D1, D2), shown here, account for 45% and 16%, respectively, of the total variance. Capital letters signify 22 different halls included in the test. Numbered vectors represent 12 different listeners. Projections of hall positions on listeners vectors approximate different listener's individual preference scales. The projection of a hall on the abscissa (D1) represents "*consensus preference*" for that hall; projection on the ordinate (D2) reflects individual preference *disparities*.

To aid interpretation, it is customary to represent the results of a factor analysis of perceptual data in geometric form, as a *perceptual space*. In our case, where the perceptual task is one of assigning subjective preferences, the perceptual space is also called a "preference space."

Figure 9 shows dimensions one and two of the four-dimensional "preference space." The capital letters represent the 22 different concert halls included in the test and the numbered vectors represent the 12 different listeners.

The meaning of the preference space is the following: The projections of the points representing the different halls on the vector of a given listener represent *that* listener's preference rating of the different halls. With more than four listeners, these data cannot necessarily be represented completely in just four dimensions. However, the mean value of the correlation coefficients, between the preference scales obtained by projecting the halls on the listener's vector and the originally determined preferences, is 0.9 for the four-dimensional solution. Thus, for all practical purposes, these projections *are* the original preferences.

What is the meaning of the rectangular coordinate axes D1 and D2? Suppose one wishes to represent the entire preference data by a single one-dimensional scale with the least rms error. In order to do this, one would have to project each hall, i.e., the positions of each capital letter, onto the abscissa D1. The foot points of these projections would then constitute the desired one-dimensional preference scale.

The component of a listener's vectors in the direction of the abscissa is the weight that that listener gave in his preference judgments to this most significant single dimension in the preference space. Thus, listeners 5, 4, and 3 have weights on Dimension 1 (D1) close to unity. (Unity corresponds to the distance between the origin and either endpoint of the abscissa or ordinate.) Listeners 9 and 7, by contrast, have comparatively small weights on D1.

Although the listeners have different weights on D1, they all have positive weights, i.e., if a hall (say F in Fig. 9) is to the right of another hall (say Y), hall F is preferred over hall Y by *all* listeners. Thus, we may call D1 the "consensus preference" factor. We will look at the possible *physical* significance of this subjective factor below.

Forcing our original data into a single one-dimensional scale does considerable violence to the data and may mean the discarding of significant information on the interplay of our listeners with the halls under study. If we allow ourselves two dimensions to represent the essentially four-dimensional data (again with minimum rms error), we get the preference space shown in Fig. 9. Since our listener's vectors have both positive *and* negative components in the direction of the ordinate (Dimension 2), we may call this dimension the "first individual-difference" factor (*first*, because Dimensions 3 and 4 also represent individual differences, albeit of smaller significance). Thus, for example, hall A is ranked very high by listener 9 but is among the least preferred by listener 7, who, incidentally, likes halls E and N best.

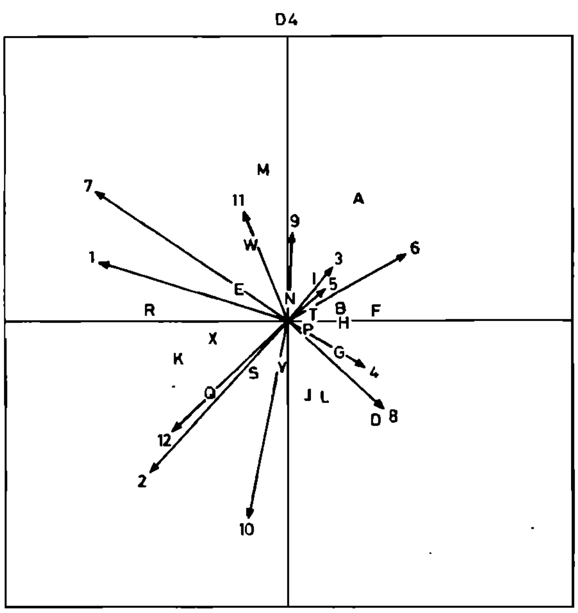

FIG. 10. Dimensions 3 and 4 (D3, D4) of the preference space accounting for 12% and 7% of the total variance, respectively, and reflecting individual differences of doubtful significance. The meanings of capital letters and numbered vectors are the same as in Fig. 9.

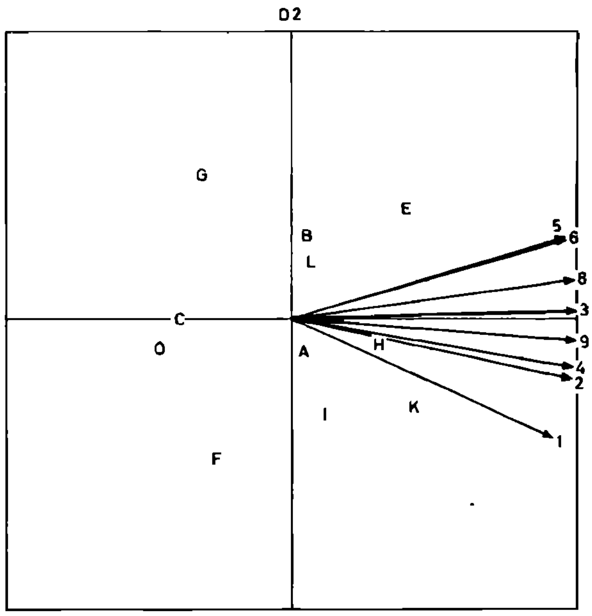

FIG. 11. Preference space for 11 halls with reverberation times up to 2.2 sec. Variance accounted for by consensus preference factor (D1): 88%.

Figure 10 shows Dimensions 3 and 4 of our four-dimensional solution. Now, as can be seen, our listeners really show their differences: their vectors point into all four quadrants of this subspace.

Since, in this study, we are *not* concerned with individual differences, we will focus our attention on Dimension 1, which we have called the consensus-preference factor. After all, concert halls are built for *many* people to listen in and we can, at best, design them so that every design feature adds something positive to the preference of a *majority* of listeners. (However, as our understanding of the underlying physical parameters improves, it is conceivable that future architects design halls which, in different seating areas, will favor different individual tastes.)

Figure 11 shows Dimensions 1 and 2 of the four-dimensional preference space for only half of our total number of halls; namely, those 11 halls that have reverberation times less than 2.2 sec. We studied these halls separately because our measurements were made in the unoccupied halls which, in the case of many of the older halls with hard wooden seats, had excessively long reverberation times, uncharacteristic of the occupied halls.

Here, Dimension 1 accounts for 88% of the total variance, and the individual preference differences are much less significant. In fact, while hall K is most preferred by listener 1, K still occupies the second-preference place for listener 5—the listener most different from listener 1.

Thus, the 11-hall data shown here seem more appropriate to our next task, namely that of correlating the "consensus preference factor" with the geometrical and acoustical parameter of these halls.

III. CORRELATION WITH OBJECTIVE PARAMETERS

Among the objective parameters we have considered are the following:

V the volume of the hall

W the width of the hall (at the first sidewall reflection in the case of nonrectangular halls)

G the time delay ("gap") between the direct sound and the first reflection at the listener's seat[7]

T the reverberation time obtained by a straight-line fit to the *first* 15 dB[8] of the decay obtained by the integrated tone-burst method

D the "definition," i.e., the energy in the first 50 msec of the impulse response divided by the total energy[9]

C the interaural "coherence" (the maximum of the cross-correlation function between the impulse responses at the two ears)

The geometric parameters *V* and *W* are obtained from drawings of the halls. The acoustic parameters *G*, *T*, *D*, *C* are obtained from measured impulse response at the dummy's ears with a powerful (5 kV) spark source on the stage.

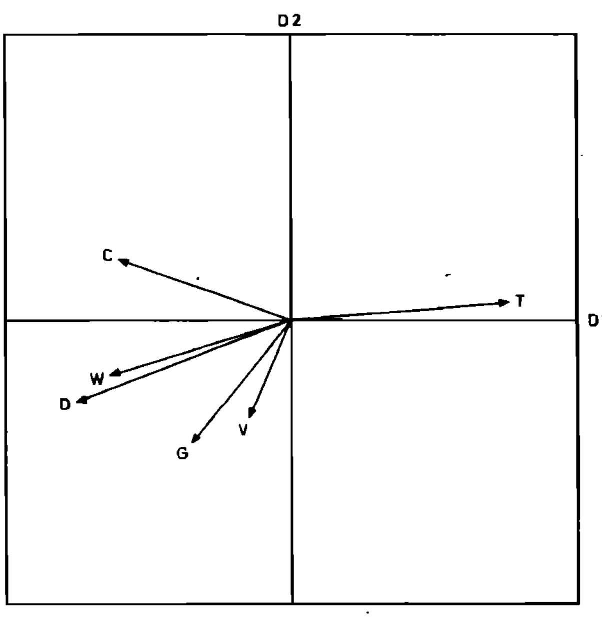

FIG. 12. Correlations of objective parameters of 11 halls with their subjective positions (cf., Ref. 11, Fig. 9) (V=volume, W=width at first side-wall reflection, T=*initial* reverberation time, D="definition," G=initial time delay gap, and C=interaural coherence; see text).

Figure 12 shows the six objective parameters as vectors in the two-dimensional preference space for the 11 halls. The component of an objective parameter vector in the direction of D1 or D2 is the correlation coefficient of that parameter with the projection of the different halls on D1 or D2. The direction of an objective-parameter vector is the direction of maximum correlation with the different halls. The vector's length indicates the magnitude of that correlation. (A correlation of 1 corresponds to the distance from origin to one of the endpoints of the abscissa or ordinate.)

According to Fig. 12, reverberation time *T* is almost coincident with D1, the consensus preference factor. Thus, the greater the reverberation time, the greater the consensus preference from these halls.

Conversely, the *definition* D (the relative energy in the first 50 msec of the impulse response) has a highly significant *negative* correlation with preference. However, this is *not* an independent datum because for these (and all other tested) halls we found a high negative correlation between reverberation time *T* and definition D: the greater the *T*, the smaller the D, and vice versa, as expected.

Another objective factor showing high (negative) correlation with D1 is the width *W* of the halls (for nonrectangular halls, *W* is the width at the first side-wall reflection). However, *W* also shows a substantial correlation (−0.62) with *T* for our 11 halls. Thus, while width may be a significant parameter (in the sense that wider halls are less preferred), we do not consider our present data as sufficient to draw a firm conclusion on hall width. But there is enough of a *suggestion* here of a possible subjective significance of a hall's width to tempt one to take a closer look.

The third parameter that shows a significant correlation with D1 is C, the coherence of the two ear signals. We found this acoustic parameter, which is a measure of how similar the two ear signals are, to be nearly uncorrelated with reverberation time *T* for the halls under study. Thus, we feel fairly certain that this parameter has an independent subjective significance, a feeling that gets further support from a corresponding analysis of the 13 halls with reverberation times above 2 sec, as shown in Fig. 13.

For these halls, the reverberation time *T* is, in fact, negatively correlated with D1, as might be expected; the greater the *T* above 2 sec, the smaller the consensus preference. But the correlation of *T* with D1 is not very pronounced. On the other hand, the correlation of *T* with dimension D2 (the first individual difference factor) is fairly large. Thus, some listeners prefer greater reverberation times and others smaller *T* values in this range of reverberation times.

The interaural coherence *C* is again almost uncorrelated with *T* and shows a strong (negative) correlation with dimension D1, so that we can again tentatively conclude that interaural coherence is a subjectively significant parameter. As before, *smaller* coherence means

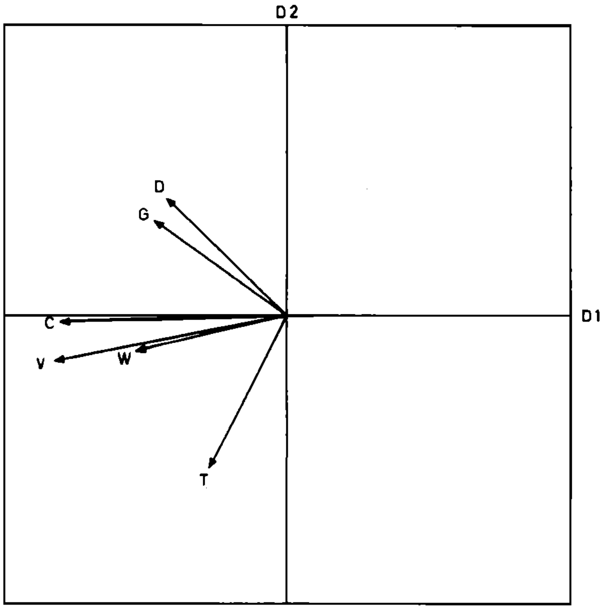

FIG. 13. Correlations of objective parameters for 13 halls with reverberation times between 2.0 and 3.2 sec. (See text for discussion of subjective–objective correlations shown in Figs. 12 and 13.)

greater preference. (This effect might be mediated by a more pronounced feeling—of being immersed in the sound—that presumably occurs for less coherent ear signals. Whether coherence is related to sound "diffusion" and whether it can be influenced by sound diffusors—on the side walls, on the ceiling, or suspended from the latter—remains to be seen.)

Interestingly enough, the volume V of the halls also becomes a significant parameter for these (generally larger) halls. The correlation with consensus preference is, in fact, the most negative of all objective parameters. The implication of this finding is fairly obvious (if economically disastrous): once a hall has reached a certain size, don't make it any bigger—or be prepared to suffer acoustically.

IV. FUTURE WORK

As mentioned above, concurrently with the recording of the musical material we also recorded the impulse from a spark source located on the stage. Thus, for each measured seat, we have available a set of binaural *impulse* responses. (In fact, it was these impulse responses from which we determined the acoustic parameters.)

However, we recorded these spark impulses for yet another reason. Not unexpectedly, reverberation time was found to be one of the significant preference factors in the halls with reverberation times below 2.2 sec. This is not really a new finding. Yet the halls at our disposal have a wide range of reverberation times. We would be much happier if we could have worked with a variety of different halls with *identical* reverberation times so that we could focus our attention better on the still *uncertain* factors affecting preference.

With the 50-odd binaural impulse responses that are already available to us in digital format, we can now, on the computer, modify these responses and filter any desired musical material with the altered responses. In particular, we can transform all responses to have the same reverberation time, say 2 sec, by simply multiplying the original impulse responses with appropriate exponential time functions! A preference analysis based on these "decay-normalized" responses should reveal additional factors which the present study could only hint at.[10]

We hope that subjective evaluations based on room responses, digitally (or otherwise) modified to control selected objective parameters, will further enhance the usefulness of our method in architectural acoustics.

ACKNOWLEDGMENTS

We thank the Deutsche Forschungsgemeinschaft for their support of this work, the Stiftung Volkswagenwerk for funding our computer, and W. Steinhoff for designing and building the periphery for acoustical signal processing.

*Based on invited talk at the 85th Meeting of the Acoustical Society of America in Boston, Massachusetts, April 1973 (J. Acoust. Soc. Am. 54, 278–279).

[1]V. Mellert, "Construction of a Dummy Head After New Measurements of Thresholds of Hearing," J. Acoust. Soc. Am. 51, 1359 (1972).
[2]P. Damaske, "Head-Related Two-Channel Stereophony with Loudspeaker Reproduction," J. Acoust. Soc. Am. 50, 1109 (1971).
[3]A. N. Burd, Rundfunktechnische Mitteilungen 13, 200 (1969).
[4]M. R. Schroeder and B. S. Atal, "Computer Simulation of Sound Transmission in Rooms," IEEE Int. Conv. Rec. 7, 150–155 (1963).
[5]K. Yamaguchi, "Multivariate Analysis of Subjective and Physical Measures of Hall Acoustics," J. Acoust. Soc. Am. 54, 1271–1279 (1972).
[6]P. Slater, "The Analysis of Personal Preferences," Br. J. Stat. Psychol. 8, 119 (1960).
[7]L. L. Beranek, *Music, Acoustics and Architecture* (Wiley, New York, 1962).
[8]B. S. Atal, M. R. Schroeder, and G. M. Sessler, "Subjective Reverberation Time and Its Relation to Sound Decay," Proc. 5th Int. Congr. Acoust. 1*b*, G32 (1965).
[9]E. Meyer and R. Thiele, "Raumakustische Untersuchungen in zahlreichen Konzertsälen und Rundfunkstudios unter Anwendung neuerer Messverfahren," Acustica 6, Akustische Beihefte, Heft 2, 425 (1956).
[10]M. R. Schroeder, "Computer Models for Concert Hall Acoustics," Am. J. Phys. 41, 461–471 (1973).