ARCHIVES OF ACOUSTICS
19, 3, 367–380 (1994)

THE FORMATION OF ONSET TRANSIENTS IN ROOMS MEASURED BY CROSSCORRELATION

J. ŻERA and M. DAROWSKA

Laboratory of Musical Acoustic Chopin Academy of Music
(00-368 Warszawa, ul. Okólnik 2)

A crosscorrelation measure for the degree of change in onset transients caused by reflections in a room was investigated. The measurements were based on a numerically modelled room. It was shown that the variability of the measure depends on the rise time of the source signal and on the delay of the measuring window from the beginning of the sound. The introduced measure is related to the early/late energy parameters of a room, and is associated with a subjective evaluation of the transients.

1. Introduction

The formation of transients in the acoustic signal appears to be an important effect of strong early reflections during the propagation of sound in rooms. Under certain conditions the change in the onset transient may be responsible for the perceived quality of the propagated sound. An acoustic transient may affect the intelligibility of speech. The intelligibility decreases when a delayed single reflection is added [13]. An acoustic transient may also affect the timbre of musical instruments. It is well known from the manipulations of onset transients in modern electronic music that onset plays a substantial role in timbre perception. A change in the onset influences the ability to identify correctly musical instruments [14]. Therefore, the nature and degree of the onset transient change in the transmission of the acoustic waveform through a room should be investigated to determine its influence on the subjective impression of the acoustic qualities of sound.

The formation of transients in the acoustic signals in rooms has received relatively little attention in the literature. The only work to which we may refer is that of CZARNECKI [6] who undertook a detailed analysis of amplitude and shape variations during the onset of sound for a relatively simple case with a limited number of sound reflections. He defined two independent sources of amplitude disturbances: "amplitude irregularity" and "phase irregularity". Czarnecki's study indicated that the amplitude irregularity depends only on the character of an impulse response of a room. This holds for signals with no specific phase across such frequency components as Gaussian noise. Phase irregularity appears for periodic signals and results from the interaction of the delays of reflected waveforms and the frequency of the propagated signal. The irregularity of the total waveform depends on the specific phases of superimposed reflections.

In the present work we are interested in how the source signal differs from the signal recorded at some observation point in a room and whether this difference is perceivable by a listener. A normalized crosscorrelation coefficient is used as a measure of this difference. The crosscorrelation function has been used primarily to study the spaciousness and diffuseness of sound fields in rooms [1–5, 11, 12, 15, 16]. These studies have focussed on the role of lateral reflections for binaural difference cues in creating the spatial impression of sound. The major finding was that the decrease of the crosscorrelation between lateral signals is related to the growth of the subjective impression of spaciousness. In the present study the crosscorrelation function is used to measure the degree to which the onset transient of a sound transmitted in a room remains similar to its origin, that is to the onset transient of the source signal. A similar approach has recently been used by CZYŻEWSKI [7] to compare the input and output signals of digital reverberators.

Measurements were made in the monophonic condition to simplify the analysis. The monophonic condition does not reflect all of the complexities of spatial impression involved in onset transient perception but is justified by the preliminary character of this investigation. Nevertheless, this investigation includes an analysis of a multiple set of reflections in three-dimensional space. This is a more realistic approach than that of CZARNECKI [6] who calculated patterns of transient buildups using the one-dimensional model.

The present research was designed to address the following questions:

a) What is the range of the predefined correlation between signals recorded at various observation and source points in a room.

b) How does this parameter depend on the impulse character of the source signal, and

c) What is the relation between the correlation coefficient and other parameters often used in room acoustics: reverberation time, early reverberation time, delay of first reflection, definition, clarity.

2. The crosscorrelation criterion for monophonic comparison of sound in a room

We seek a simple numerical parameter based on the CrossCorrelation function which would represent the similarity between the onset transient of the source waveform and the onset transient of the waveform received in a room. The source and the observation points are both located somewhere inside a room. This parameter, called the Maximum CrossCorrelation (MCC), is defined as the maximum absolute value of the crosscorrelation function determined over a specific observation time T:

$$\mathrm{MCC} = \max |R_{\mathrm{sr}}(\tau)|,$$

where

$$0 < \tau < T$$

$$R_{\mathrm{sr}}(\tau) = \int_{t}^{t+\Delta t} p_s(t) p_r(t+\tau)\, dt$$

MCC — maximum absolute value of the crosscorrelation function, $p_s(t)$ — source signal, $p_r(t)$ — signal recorder in a room, Δt — averaging time, $\langle 0, T \rangle$ — interval over which the crosscorrelation function is defined.

By calculating the maximum value of a correlation function, we can ignore the time delay between the signals p_s and p_r resulting from the distance between their respective locations. The MCC is simply a measure of the similarity between the signals $p_s(t)$ and $p_r(t)$ because they are compared with the relative delay revealing their best similarity.

The crosscorrelation function is, in a simple way, related to the autocorrelation function of the source signal because a recorder signal is obtained by convolving the source signal with the impulse response of a room. This relation may be used as an alternative way to calculate the crosscorrelation function. The MCC depends both upon the impulse response of a room and the source signal. In this study we concentrate on the investigation of the variability of the MCC for simple source signals, sinusoids of various frequencies.

3. Testing material

The calculation of the acoustical parameters and the synthesis of stimuli for the listening tests were based on a numerically modelled concert hall. In the model we adopted a geometrical shape and wall absorption coefficients similar to those existing in the chamber-music concert hall at the Chopin Academy of Music in Warsaw whose volume is about 800 m^3 (Fig. 1a).

The ray tracing method was used to obtain the impulse response of the room (KULOWSKI [9]). The impulse response of the room was calculated for octave-frequency bands at 250, 500, and 1000 Hz. Four source positions on the stage and twenty observation points in the audience space were investigated (Fig. 1b). We assumed that eighty echograms ($4 \times 20 = 80$) constituted a sufficient statistical representation of the possible reflection patterns in the modelled room. Each observation point was represented by a sphere of 0.5-m radius. For each combination of source and observation point, a total of 32.000 rays were emitted in random directions with equal probability to calculate the impulse response. Each ray was traced until the energy level became — 40 dB re its initial-value.

In typical application the impulse response of the room is obtained by cumulating energy arriving at the observation point in predefined time bins. This calculation, characteristic of a ray tracing method, is especially designed for calculating the energy of the room (e.g., reverberation time or energy arriving in specified time intervals) because the phase relations between reflected sounds are not preserved. The method is not sufficient, however, when it is necessary to obtain an amplitude response in which exact delays between wavefronts approaching the observer should be preserved. Those are essential to compute correctly the signal waveform at the observation point.

The following procedure was used to obtain correct estimation of the amplitude impulse response without forfeiting the phase relations. Each group of N rays moving in the same general direction, that is reflecting from the same set of walls, was separated. Then, this group was treated as representing a single equiphase surface propagating in a room. The energy associated with this surface was calculated as $E = \sum_1^N E_i$ (where E_i is the energy represented by a single ray) and its mean time of moving from the source to the observer as $\tau = (\sum_1^N \tau_i)/N$. Finally, the amplitude of a single peak in the amplitude impulse response was calculated as $A = \sqrt{E}$, and it was located at the time delay of τ.

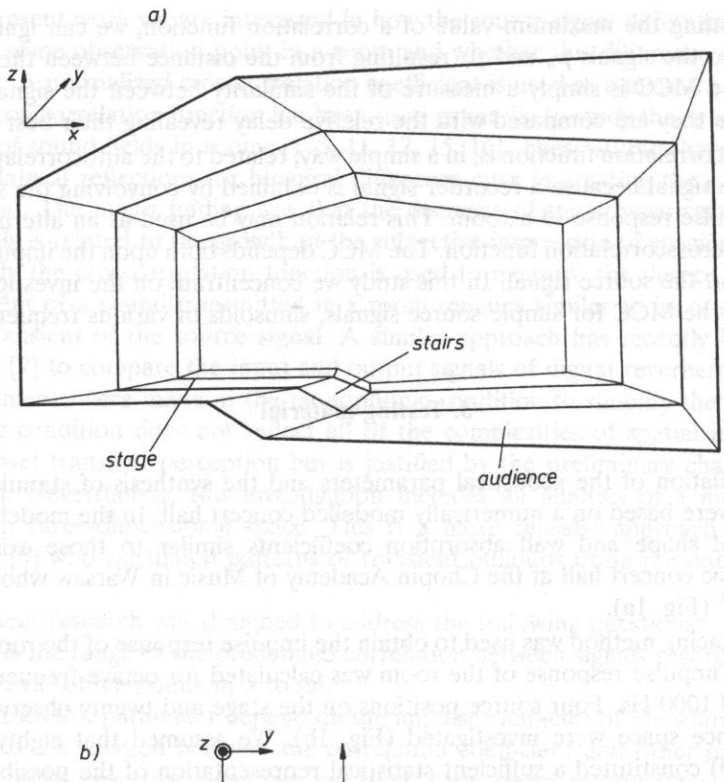

FIG. 1.

a) A three-dimensional view of the modelled room.
b) Horizontal view. Sound sources are positioned 1.2 m above the stage floor.
Observation points are 1.3 m above the audience floor.

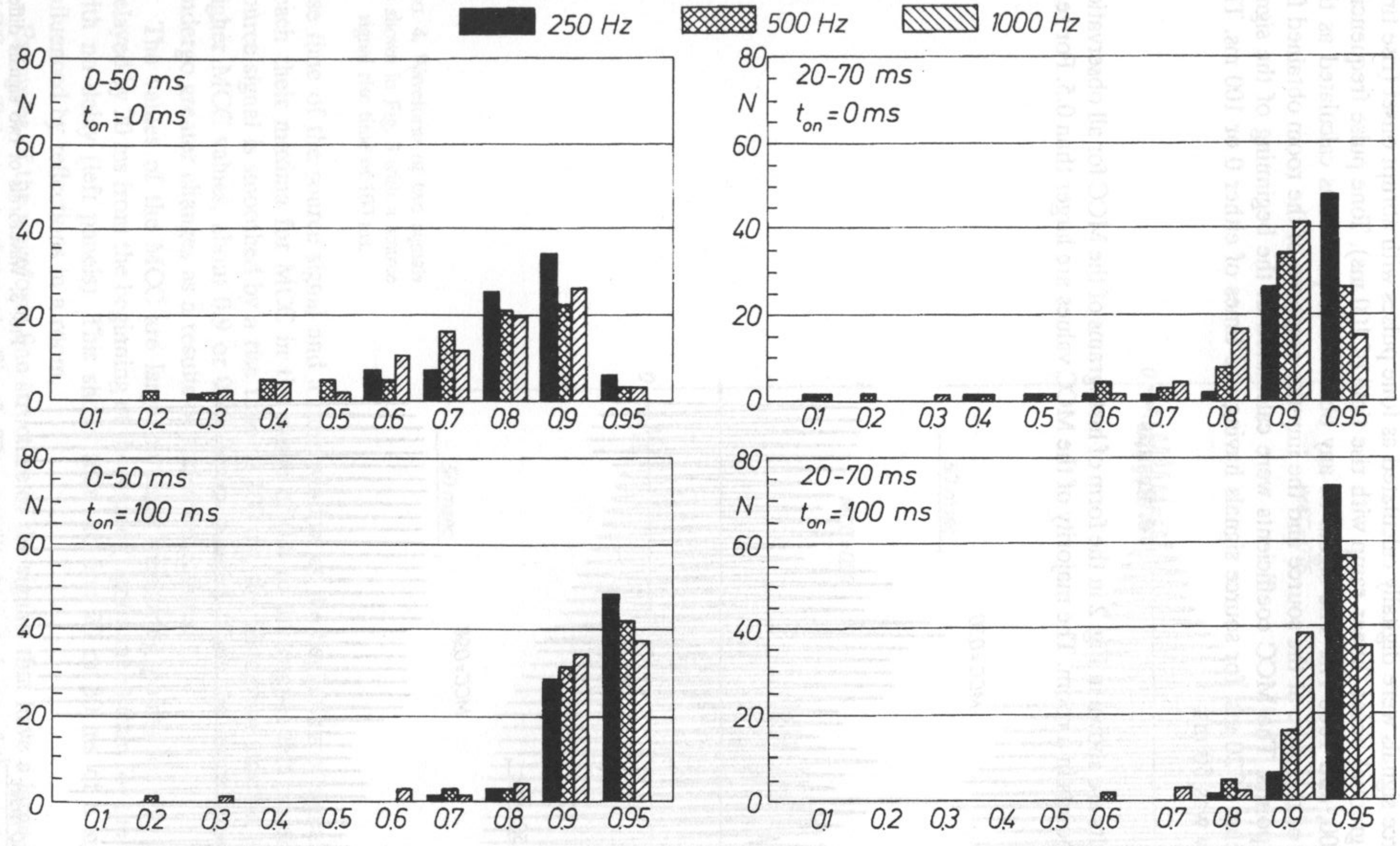

FIG. 2. Histograms of the Maximum CrossCorrelation coefficients (MCC) for pairs of source signals and signals received at observation points. The MCC values are calculated for the intervals 0–50 ms and 20–70 ms from the beginning of the sound. The source-signal rise time, T_{on}, is 0 or 100 ms. The source-signal frequency is 250, 500, or 1000 Hz.

The source signals were digitally calculated as tone pulses with abrupt onset (rise time of 0 ms) or gradual onset (linear ramp with rise time of 100 ms). Tone pulse frequencies were 250, 500, and 1000 Hz. The signal at any observation point was calculated as the convolution of the wave at the source and the impulse response of the room obtained for this pair of points. The MCC coefficients were calculated at the beginning of the signal or with a delay of 20 ms, for source signals having rise times of either 0 or 100 ms. The time window was 50 ms.

4. Results

The results are shown in Fig. 2 in the form of histograms of the MCC for all observation and source points in a room. The majority of the MCC values are larger than 0.5. For zero

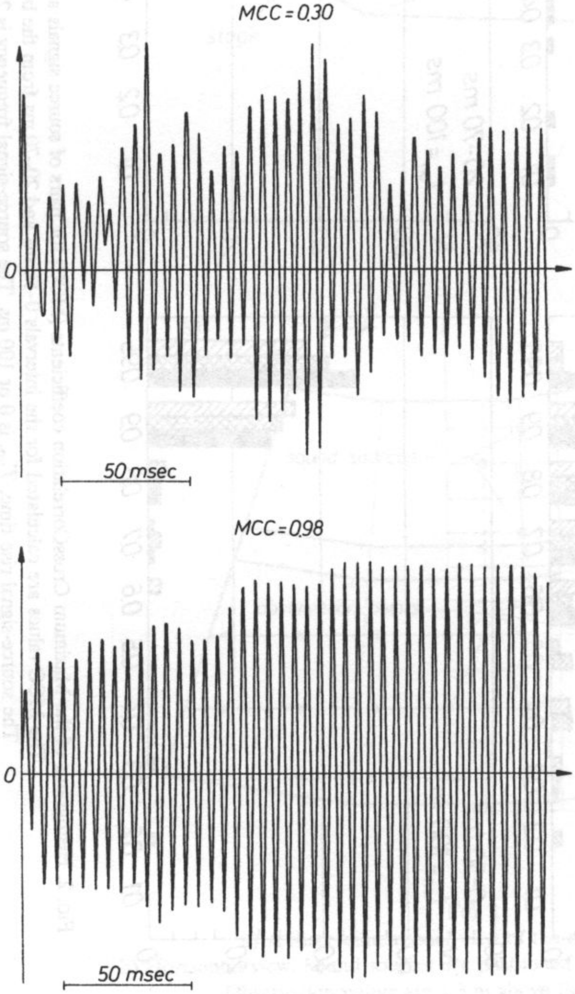

FIG. 3. Waveforms of two signals differing in MCC values. The rise time of the source signal is 0 ms.

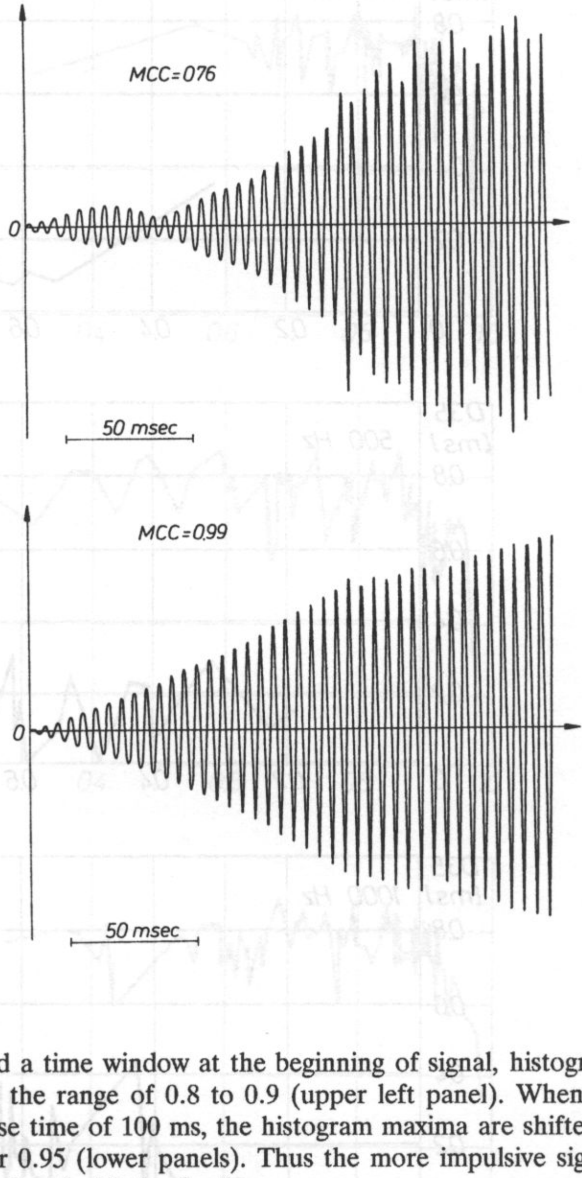

FIG. 4. Waveform of two signals as shown in Fig. 3 with a source signal rise time of 100 ms.

rise time of the source signal and a time window at the beginning of signal, histograms reach their maxima for MCC in the range of 0.8 to 0.9 (upper left panel). When the source signal is smoothed by a rise time of 100 ms, the histogram maxima are shifted to higher MCC values, about 0.9 or 0.95 (lower panels). Thus the more impulsive signals undergo greater changes as a results of added reflections.

The values of the MCC are larger at most measuring points for the time window delayed by 20 ms from the beginning of the sound (right panels) than for the time window with no delay (left panels). This shows that the first 20 or 30 ms of a signal is most influenced by reflections in a room.

Examples of the waveform fine structure of the signals that have a large difference in MCC coefficients are shown in Fig. 3. The amplitude envelope of the signal with MCC = 0.98 is relatively smooth and rises gradually to its steady state value (lower panel).

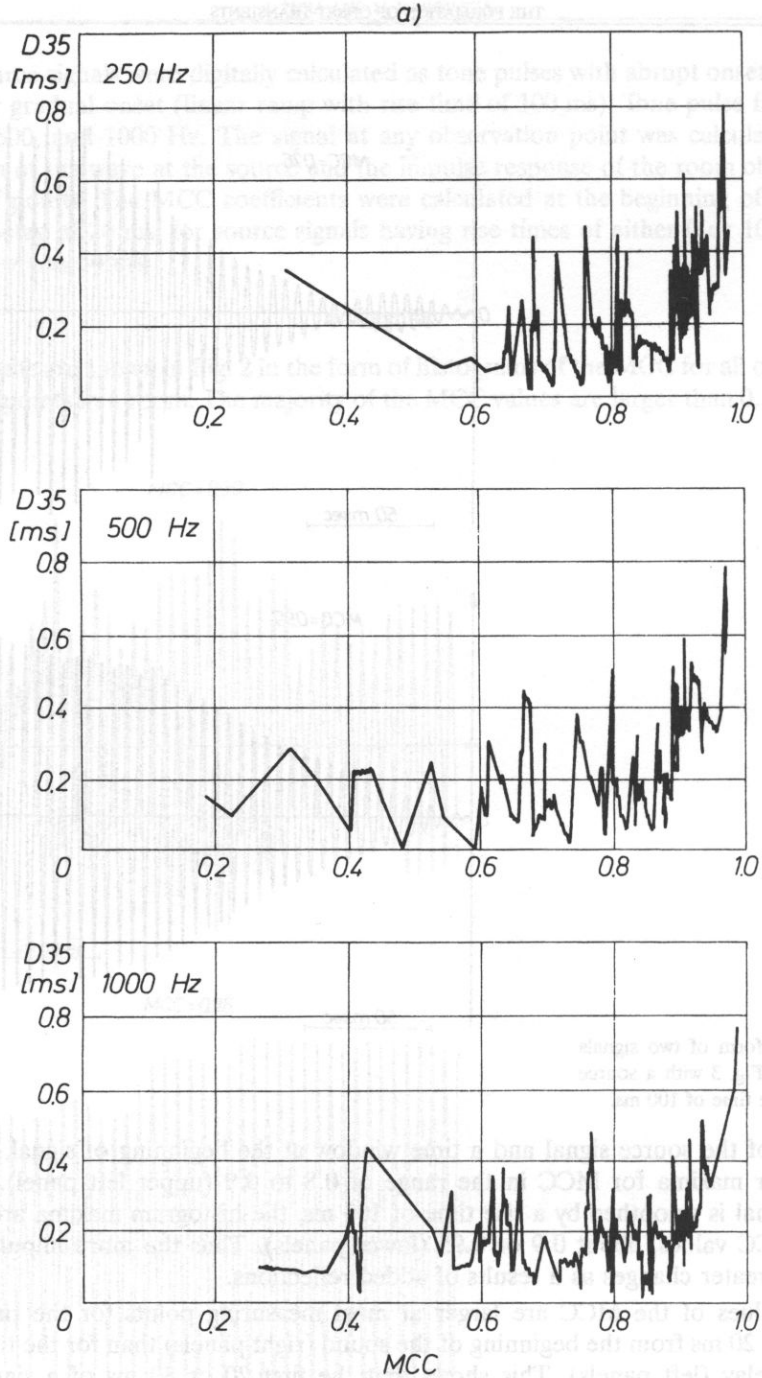

FIG. 5a. The relation between the MCC coefficient and the ratio of energy arriving in the first 35 ms, D35, of the echogram for the three signal frequencies.

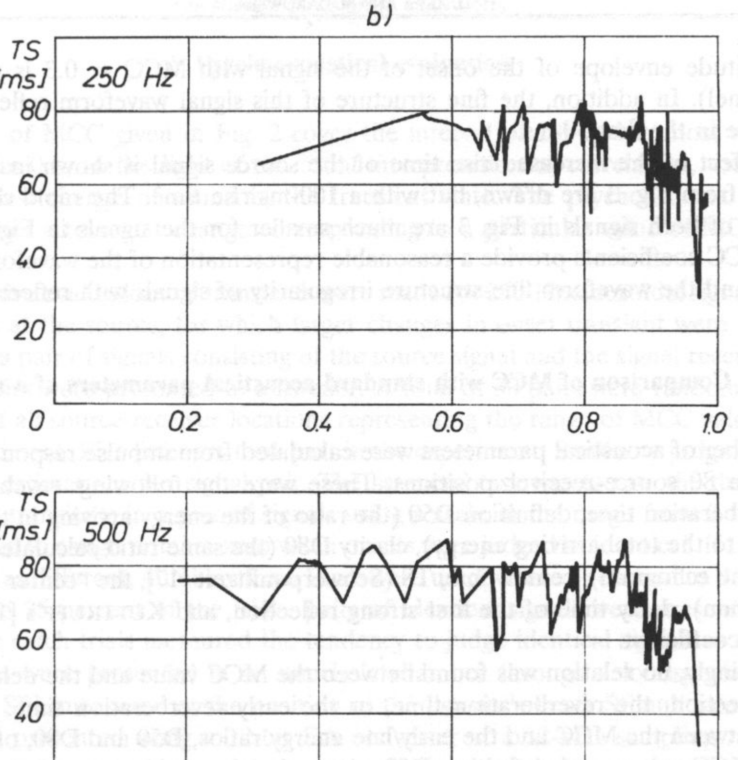

FIG. 5b. The relation between the MCC coefficient and center time, TS, of the echogram.

The amplitude envelope of the onset of the signal with MCC = 0.3 is very irregular (upper panel). In addition, the fine structure of this signal waveform reflects the phase disturbance in the first 50 ms.

The effect of the increased rise time of the source signal is shown in Fig. 4, where the signal from Fig. 3 are drawn, but with a 100-ms rise time. The rapid changes in the amplitude of both signals in Fig. 3 are much smaller for the signals in Fig. 4. Summing up, the MCC coefficients provide a reasonable representation of the waveform amplitude envelope and the waveform fine-structure irregularity of signals with reflections added.

5. Comparison of MCC with standard acoustical parameters of a room

A number of acoustical parameters were calculated from impulse responses of a room taken from 80 source-receiver positions. These were the following: reveberation time, early reverberation time, definition D50 (the ratio of the energy arriving in the first 50 ms of a signal to the total arriving energy), clarity D80 (the same ratio calculated for the first 80 ms of the echogram), center time, TS (Schwerpunktzeit [17], the "center of gravity" of the echogram), delay time of the first strong reflection, and KUTTRUFF's [10] "temporal diffusion" coefficient.

Surprisingly, no relation was found between the MCC value and the delay of the first strong reflection, the reverberation time, or the early reverberation time. There was a relation between the MCC and the early/late energy ratios, D50 and D80, or center time, TS. The MCC values and definition D50 were related in such a way that for values of MCC larger than 0.9 an increase of D50 was also observed. A similar effect was less apparent for clarity D80. Furthermore, this tendency was much clearer when the definition was calculated for the first 35 ms of the echogram, rather than for the first 50 ms. We refer to this parameter as to the definition-D35, by analogy to definition-D50. The results were less clear when the interval over which the early energy was cumulated became smaller than 35 ms. This results confirms the previous conclusion, based on histograms shown in Fig. 2, that the first 30 ms of the signal is most affected by reflections in a room.

The MCC values versus the definition D35 are shown in the left panels of Fig. 5, for source signals having a 0-ms rise time. It is evident that large values of D35 are associated with large values of MCC. The D35 increases from about 0.2 or less to values larger than 0.4 when the MCC becomes larger than 0.9, that is for MCC values representing a little change of transients. The opposite tendency is observed for TS values (right panels of Fig. 5). The TS becomes smaller for the transients less influenced by reflections (large MCCs). This decrease of TS can be predicted from the D35 change because TS and early/late energy ratios are not independent parameters. The growth of the amount of energy in the first 35 ms of the echogram causes a shift of the TS to the beginning of the echogram.

The relation between D35 or TS and the MCC is of importance because D35 (usually calculated as D50) and TS are parameters applied in room acoustic. Thus the present results show that these standard parameters are useful in estimating the size of the change in the onset of the signal after a transmission through the room. Only a small change in transients should be expected when the value of D35 is greater than 0.4.

6. Psychoacoustical evaluation

The values of MCC given in Fig. 2 cover the interval of 2.0 to 1.0, most values being greater than 0.8. In the light of the usual interpretation of correlation, these values suggest a modest change of sound in the transient state. The psychoacoustic data will establish whether a transient change corresponding to a particular value of MCC is perceptible.

The subjective evaluations of transients were made for 250-Hz sinusoidal signals having 0-ms rise time at the source, for which larger changes in onset transient were observed. On each trial, a pair of signals consisting of the source signal and the signal received at an observation point were presented to a listener. A total of 30 pairs were randomly chosen from the set of all source-receiver locations representing the range of MCC values in the previous experiment. Five listeners took part in the experiment. Similarity judgments were based on four estimations per signal pair. The listeners' task was to rate the dissimilarity between onset transients in a pair of signals using a scale in the range from 0 to 6. They were introduced to assign number zero when onsets were judged as identical. If the signals were regarded as different, numbers from 1 to 6 were used, where 6 referred to the greatest dissimilarity. In 25 percent of the trials, pairs of identical signals were presented to the listener. These catch trials measured the tendency to judge identical signals as different.

The sounds were presented in an anechoic chamber through a loudspeaker (Alton 140) at 60 dB SPL measured at the position of the listener's head. Stimuli were digitally calculated and presented using a 12-bit DA converter at a 32-kHz sampling rate. Each pair of signals was presented three times to the listeners before they responded.

The results of subjective measurements of dissimilarity ratings for transients having different values of MCC are shown in Fig. 6. The values are means taken over three subjects. It is clear from Fig. 6 that lower dissimilarity ratings are associated with larger values of MCC. Then, MCC appears to be a measure which is also related to the perception of changes in onset transients due to sound propagation in a room.

Some dissimilarity ratings, however, do not decrease with increasing MCC. These points are depicted as the small "B" area in Fig. 6. The signal pairs were judged as different despite the associated large value of MCC, indicating that the correlation coefficient based on the fine structure of signal waveforms was not an appropriate measure for these pairs of signals when one wishes to predict a listener's subjective. Presumably, the correlation is a major factor related to the perceptual similarity of signals, but it may not be the only important parameter. It may be assumed that our perceptual impression of transients has a number of dimensions that are related to additional signal parameters. Among them, the shape of the signal amplitude envelope is probably of primary importance. The signals with response shown as squares in Fig. 6 had many strong irregularities in their amplitude envelopes. All of them were judged as very dissimilar from the source signals. Those belonging to the "B" region had irregular amplitude envelopes; however, their fine structures were well correlated with the fine structures of the source signals (resulting in large MCCs). Thus the fine structure cue was not used by the listeners for the "B" region. It is possible that the envelope cue was used.

Finally, it should be stressed that pairs of signals with the highest correlation (MCC > 0.99) were clearly recognized as different by the listeners. There is no critical value of MCC in Fig. 6 above which the stimuli taken at observation points would become

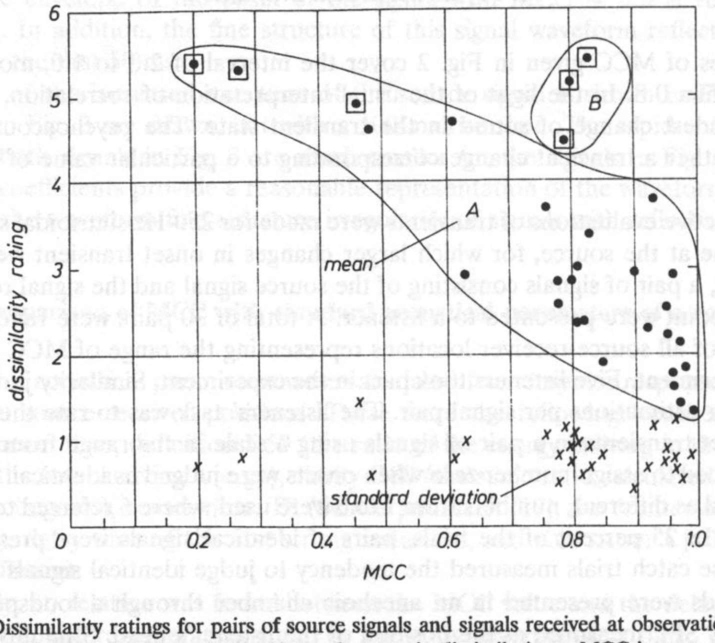

FIG. 6. Dissimilarity ratings for pairs of source signals and signals received at observation points. "A" — the set of results showing a decrease in dissimilarity as MCC increases. "B" — points showing no such relation. Circles within boxes show the results for stimuli of with onsets having many irregularities. The stimulus frequency is 250 Hz.

indistinguishable from the source signals. Due to the fact that all signals in the control catch trials were correctly judged as identical, it is clear that the subject's responses were not biased toward a judgement.

7. Conclusions

It was shown that the maximum value of the crosscorrelation function (MCC) between the source waveform and waveform corresponding to an observation point in a room calculated over the first 50 ms of signal durations is a parameter that describes well transient changes due to reflections in a room. This parameter is sufficiently sensitive to differentiate numerically between pairs of signals at different locations in a room. In our model of a room MCC values ranged from 0.2 to 1.0.

Relations of MCC to two well recognized parameters in room acoustic, the definition D35 and TS, were obtained. These relations predict that for values of D35 greater than 0.5, reflections will have a very small influence on onset transients.

Finally, dissimilarity ratings from human observers show that the MCC parameter is related to the subjective evaluation of the transients. The ear appears to be rather sensitive to transient build up associated with room reflections, as the transients were easily distinguished from their sources even for MCC larger than 0.99.

These conclusions should be treated with some caution as the data presented in this work were calculated for only one model room, a medium size chamber concert hall. This model hall corresponds to an existing hall with good acoustical characteristics. Rooms of different sizes and having different subjective evaluations are to be examined in future work.

As a final note, some musical interpretation may be made basing on the obtained results. The data shown in Fig. 2 suggested that the influence of reflections is the strongest during approximately the first 30 ms of a signal and for signals having short onset times. The sounds of many instruments have onset times shorter than 50 ms. It has been reported [8] that guitars subjectively rated as having superior sound quality have rise times in the range of 30 to 70 ms, with quality ratings diminishing for shorter rise times (e.g., less than 20 ms). This would be consistent with the notion that the sound produced by instruments judged as having superior sound quality are less disturbed by room reflections than the sound of instruments judged as having poor sound quality. Quality ratings for the sound of such instruments should be less dependent on the sound field in a room, including the listener and performer position.

Acknowledgments

The authors are very indebted Dr Andrzej KULOWSKI for his computer program of the ray tracing method, which he made available. We thank Mr. David EDDINS, Dr Antoni Jaroszewski and Dr Beverly WRIGHT for their comments on the earlier draft. We also thank Dr Witold STRASZEWICZ for his helpful criticism and discussions. This work was supported by the grant number CPBP 02.03.VIII.8.2 from the Polish Academy of Sciences.

References

[1] Y. ANDO, *Subjective preference in relation to objective parameters of music sound fields with a single echo*, J. Acoust. Soc. Am. **62**, 1436–1441 (1977).

[2] Y. ANDO, K. KAGAYAMA, *Subjective preference of sound with a single early reflection*, Acoustica **37**, 111–117 (1977).

[3] Y. ANDO, D. GOTTLOB, *Effects of early multiple reflections on subjective preference judgements of music sound fields*, J. Acoust. Soc. Am. **65**, 524–527 (1979).

[4] Y. ANDO, M. IMMAMURA, *Subjective preference tests for sound fields in concert halls simulated by the aid of computer*, J. Sound Vib. **65**, 228–239 (1979).

[5] Y. ANDO, Y. KURIHARA, *Nonlinear response in evaluating the subjective diffuseness of sound fields*, J. Acoust. Soc. Am. **80**, 833–836 (1986).

[6] S. CZARNECKI, *Irregularities in acoustic signals in rooms*, Doctoral thesis, Institute of Fundamental Technological Research, Polish Academy of Sciences, (1958) (in Polish).

[7] A. CZYŻEWSKI, *A method of artificial reverberation quality testing*, J. Audio End. Soc. **38**, 129–140 (1990).

[8] A. JAROSZEWSKI, A. RAKOWSKI, J. ŻERA, *Opening transients and the quality of classic guitars*, Archives of Acoustic **3**, 79–84 (1978).

[9] A. KULOWSKI, *Computer modelling of acoustic field in rooms*, Scientific report CPBP 02.03. VIII.8.6 (1989) (in Polish).

[10] H. KUTTRUFF, *Uber Aotukorrelationmessungen in der Raumakustik*, Acustica **16**, 166–174 (1965/66).

[11] J. P. A. Lochner, J. F. Burger, *The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of sounds*, Acustica **8**, 1–10 (1958).
[12] R. W. Muncley, A. F. B. Nickson, P. Dubont, *The acceptability of speech and music with a single artifical echo*, Acustica **3**, 168–173 (1953).
[13] T. Nakajima and Y. Ando, *Effects of a single reflection with varied horizontal angle and time delay on speech intelligibility*, J. Acoust. Soc. Am. **90**, 3173–3179 (1991).
[14] P. Schaeffer, *Traité des objets musicaux*, Editions du Seuil, Paris 1974.
[15] M. R. Schroeder, D. Gottlob, K. F. Siebrasse, *Comparativestudy of European concert halls: correlation of subjective preference with geometric and acoustic parameters*, J. Acoust. Soc. Am. **56**, 1195–1201 (1974).
[16] H. Yanagawa, Y. Yamasaki and T. Itow, *Effect of transient signal length on cross-correlation functions in a room*, J. Soc. Acoust. Am. **84**, 1728–1733 (1988).
[17] M. Vorlander, *Ein Strahlverfolgungs-verfahren zur Berechnung von Schallfeldern in Raumen*, Acustica **65**, 138–148 (1988).