Proceedings of The Institute of Acoustics

DEVELOPMENT OF A COMPUTER-BASED BUILDING ACOUSTICS SIMULATOR

P T LEWIS and C B CHINOY

WELSH SCHOOL OF ARCHITECTURE UWIST

1. INTRODUCTION

Existing methods of rating the capacity of walls and partitions to provide speech privacy are based on how well they reduce steady state noise. However, there is ample evidence from auditorium acoustics that the intelligibility of speech does not simply depend on its amplitude (compared with the level of background noise) but also on the time pattern of energy arrival for individual speech sounds. Indeed, much of the effort in the design of auditoria goes into achieving a satisfactory performance in the time domain, in terms of early reflected energy and an appropriate reverberation time.

The question therefore arises that, if this is true for the intelligibility of speech heard in auditoria, can and do the temporal characteristics of the energy arrival also affect the intelligibility of speech when it is transmitted through building structures.

The facility described in this paper was developed to allow this idea to be evaluated. It involves simulation of the process of sound transmission through buildings using a computer and thereby gives great flexibility and control over the range and properties of the situations that can be examined.

2. GENERAL PRINCIPLES

The facility had to satisfy two main requirements:

1. to be able to capture, describe and modify the temporal characteristics of a given transmission situation

2. to simulate how speech (or any other) signal would be heard if it had been transmitted along this path

The basis of the method adopted was that, for any time-invariant linear system, the output signal y(t) is related to the input signal x(t) by the impulse response h(t) and the convolution integral:

$$y(t) = \int_{-\infty}^{t} x(\tau)\, h(t-\tau)\, d\tau$$

or, in the time domain:

$$Y(i\omega) = X(i\omega).H(i\omega)$$

where $X(i\omega)$ abd $Y(i\omega)$ are the complex) spectra of input and output signals respectively and $H(i\omega)$ is the transfer function.

In the present case, $x(t)$ represents the anechoic speech emitted at some point in the source room, $y(t)$ is the speech signal heard at some point in the receiving room and $h(t)$ is the combined impulse response of the transmission paths between the two points.

The authors had previously developed a method for capturing an impulse response in digital form using proprietary digital signal processing components and a PDP 11/05 minicomputer. However, this system was limited to signals that would fit into the memory of the computer i.e. to signals with a maximum length of the order of 1.2 secs sampling at a rate of 20KHz. The simulation facility needed to be able to handle signals very much longer than this and to be able to perform digital to analogue conversion as well as analogue to digital conversion. The new technique is therefore based on our original work but, because of these increased requirements, has received substantial modifications and additions.

3. DESCRIPTION

A block diagram of the equipment is given in Figure 1. Its operation takes place in four stages.

STAGE 1 Analogue to digital conversion

The analogue signal is fed, either directly or from a tape recorder, into the analogue to digital side of the digital audio processor. The resolution used is 14 bits and the sampling frequency is variable up to a maximum of approximately 50KHz. This is the highest speed at which the general purpose interface used is able to operate and new device drivers had to be written to achieve these rates. The digitised signal is then transferred directly onto magnetic tape using the memory of the PDP 11 to provide buffering. This tape is computer compatable and is either used directly in the convolution stage or is transferred to disc for editing etc.

STAGE 2 Convolution

The digital convolution required considerable memory size and processing time and could be carried out on the minicomputer. This process therefore takes place on the University's main computer, a VAX 11/780. Convolution within the time domain is not feasible due to the length of processing time involved and this process is therefore carried the frequency domain. At present, the speech signal is Fourier and inverse-Fourier transformed using a 32K array size although a 64K array size would be preferable. These very large FFT's are required because of the length of the impulse response. Thus, if the minimum useful sampling rate of 20KHz is used, an impulse response of 0.8sec produces a digital file of 16K length. The way the convolution has to be carried out requires that this array size be doubled during the processing hence giving rise to the 32K FFT that is used. The result of the convolution integration is a digital file representing the output that would be produced by transmission along a path with the impulse response used.

STAGE 3 Digital to analogue conversion

The digital file of 'transmitted' speech is played through the D/A side of the audio processor and (generally) stored on a conventional analogue tape recorder. Ideally, of course, a digital audio tape recorder would be used but the format used for such recorders cannot readily be adapted for reading/writing with a computer.

STAGE 4 Playback of simulated transmitted speech

The final state is the replay of the processed speech signal for listening tests. This is carried out in an anechoic room using a conventional amplifier and loudspeakers.

4. CURRENT APPLICATIONS

Work on speech intelligibility in auditoria indicates that, for impulse responses of the type found within such rooms, the ear/brain divides the signal into useful energy which arrives within some 50msec of the direct sound and enhances intelligibility, and masking energy which arrives later than 50msec after the direct sound.

Previous work by the authors has shown that there are substantial differences between walls in relation to the proportion of these 'early' and 'late' energy fractions. The simulator is being used to test:

(a) whether the ear processes signals transmitted through structures in the same way that it processes speech heard in auditora, i.e. does the 50msec interval also apply in the case of the type of impulse responses encountered in structure-borne transmission and, if not, what are the integration characteristics in this case

(b) what effect this has on the intelligibility of speech heard through walls with the same steady stage performance but for which the ratio of early to late energy is different.

FIGURE 1 Block diagram of equipment