Development of Time Domain Audio System

Akira Nishikawa
Isao Yokoyama
Katsuhiko Tsumori
Hiroshi Kowaki

GOOD DESIGN AWARD

Abstract

Fujitsu Ten made a decision to develop and market a full-fledged home audio system under its own brand, in order to improve its audio technology capability and strengthen its brand image. To thoroughly differentiate such system from those of other companies, we implemented the product development based on the "Time Domain Theory" advocated by Hiroyuki Yoshii, President of the Time Domain Corporation. The feature of this theory is that it stresses the time-base characteristics of the speakers, as opposed to most conventional audio systems which stress frequency characteristics and were developed with a focus on reproducing all sounds from low to high as evenly as possible, and at reducing distortion. Specifically, this theory represents the thinking that faithful reproduction of the input waveforms is the ideal for reproduction of sound. This paper describes the content of the Time Domain Audio System that was developed.

1 Introduction

Product development by our company formerly centered on on-board equipment for automobiles, but we decided to develop and launch a full-fledged home audio system under our own brand with the aim of strengthening our brand image and enhancing our technological capability.

This system has been realized with the technological cooperation of the Time Domain Corporation (a venture enterprise based in Nara) and is a product based on the "time domain theory" advocated by its President Hiroyuki Yoshii.

President Yoshii was formerly employed in the Development Division of the Onkyo Corporation where he was involved in developing the GRAND SEPTER (GS-1), which met with a hugely favorable rating in the audio world. Ever since he has been an audio world celebrity.

2 About the "time domain theory"

Most conventional audio systems stress frequency characteristics and were developed with a focus on reproducing all sounds from low to high as evenly as possible, and at reducing distortion, whereas the feature of the time domain theory is that it stresses the time-base characteristics of the speakers. There have been theories and products focusing on time for some time now, giving attention to the time-base and to unifying the time taken for the sounds to reach the ear from the speakers, but the time domain theory takes this approach much further, representing the thinking that faithful reproduction of the input waveforms is the ideal for reproduction of sound.

The relationship between the speaker's frequency characteristics and time-base characteristics is one of Fourier transformations or reverse Fourier transformations. The frequency characteristics are of 2 types: acoustic pressure frequency characteristics and phase frequency characteristics. If the phase frequency characteristics are different, the sound will sound different to the ear even if the acoustic pressure frequency characteristics are flat. This is the reason why the although there exist many speakers that make the acoustic pressure frequency characteristics approach being flat, each produces sounds that are very different from the others. The idea of the time domain theory is that in order to reproduce sound faithfully, the ideal is that the sounds of the various frequencies produced in response to impulse input should emitted in the shortest possible time with identical timing and level; in other words that the impulses of input should be faithfully reproduced. The Time Domain Company's "Yoshii 9" cylindrical speaker launched on the market in May 2000 was developed based on this theory (refer to Fig. 1).

Fig.1 The "Yoshii 9"

3 The form that speakers should take

Roughly speaking there are two main approaches to the production of sound by speakers. One holds that "a speaker is a musical instrument", while the other takes the view that "a speaker is a transducer", that is, it is simply a converter that does not itself color the sound in any way. From this one could conclude that the answer to the question what form speakers should take is simply "the form that will give the listener enjoyment".

As is generally recognized, sound is a question of taste. Food varies from restaurant to restaurant, having a different individuality at each, and people's tastes vary widely from those who appreciate food prepared carefully from ingredients to those who, like children, find fast foods delicious. Exactly the same thing can be said about audio. Metaphorically comparing audio to food, we could call the music recorded onto a CD or other medium (or alternatively the conditions of the recording) the "ingredients" (or their quality), the processing of the reproduced sound by equalizers and surrounds the "flavorings", etc., and the selection and location of the reproducing equipment in a room to adjust the sound the "cooking".

Just as the food culture of each country is different, so does taste for sounds. It depends to a large extent on the experiences with an individual's contact with music from birth onward. For instance, people who have themselves long been pianists, and are aware powers of expression and pay careful attention to the feel of keystrokes will probably feel that an on-mike recording which has little reverberation and so makes it easier to pick out the individual sounds and facilitates writing the notes down is preferable to a recording with a lot of reverberation that obscures the feel of the keystrokes. Other people who go often to listen to music in concert halls will prefer recordings rich in warm reverberations that give an impression of physically enveloping the listener, while some others still who are violinist or cellists and are used to hearing stimulating sound close to their ear will favor recordings in which the microphone is positioned closer than normal and base sounds are emphasized via an equalizer, producing a vivid effect. And it seems that there are some audio fans who feel discomfort when they hear music from a speaker other than the one they are used to listening to.

Thus there are many different things that people want from audio, and we believe that speakers should produce sounds meeting the needs of each. To consider such tastes in sound we will take a look over the transfer functions for reproduced sound so as to examine just what sounds are actually reproduced by speakers. Fig. 2 gives an approximate illustration of the transfer functions that are involved from the emission of sound from a musical instrument up to the sound's being heard by a listener from a CD. If the aim here is to reproduce the original sound faithfully, then the following equation expresses the ideal for reproduction:

$M(f)R(f)Pl(f)A(f)S(f)R(f)=1$.

Among these functions, those principally relevant for recording are M(f)R(f). The sound sent forth by the musical instrument passes through the space in the hall, etc., into a microphone, which edits it before it is recorded by a recording machine. What should be noted here is that the sound that is recorded contains the tendencies of many different sounds. This is primarily a question of the position of the microphone. The recording engineers utilize their expertise to implement a recording that they consider ideal, but there are various ways of positioning the microphone. One method is to suspend it from the ceiling, so that the instrument's sounds and the hall's reverberations are recorded together; another is to position the mike directly beside the instrument so that only the instrument's reverberations are recorded; sometimes a method whereby electronic reverberations are added afterward is used. Furthermore if the tendencies of the sounds from the monitoring speaker are different, there will naturally be a myriad different sounds produced after the sound is subjected to equalizing using judgement by ear. Thus it is not necessarily the case that M(f)R(f) will have a value close to 1.

Next we turn to the reproducing functions P1(f)A(f)S(f)R(f). These transfer functions are generally arranged in the order of their closeness to 1, giving the order P1(f) A(f) S(f) R(f). In most listening environments, especially ordinary homes, it is impossible to make transfer functions equal precisely 1 and thus is unrealistic to aim for 1 for all of the transfer functions. Moreover there is a certain degree of fixed reverberation (preventing the transfer factors from equaling 1) in monitor speakers used in recording; when speakers with transfer functions close to 1 (without reverberation) reproduce software produced via recording/editing using monitoring speakers, reverberation may well be too low, so that the reproduction sounds stark. Amid these various considerations we held that aiming for the speaker transfer functions to equal 1 would be the rational direction to pursue, because: (a) as reverberation increases (transfer functions move further away from 1), delicate signals become masked to an equal extent and thus become difficult to reproduce; (b) it is almost impossible to completely restore a transfer function once it has become deformed (that is, return back to 1 a transfer function that has strayed from 1), especially in speaker reproduction; and (c) the transfer functions of only a

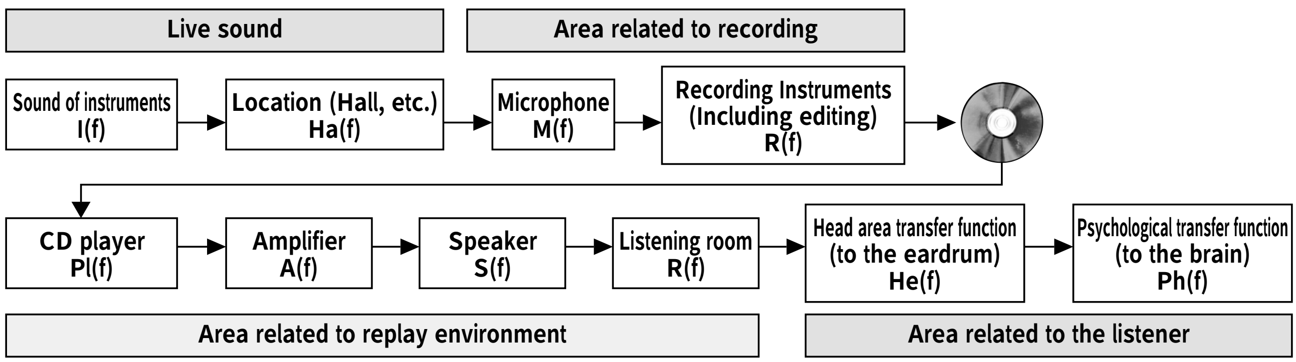

Fig.2 Transfer functions from emission of sound up to perception by humans

moment ago will all become simple during reproduction (that is, after sounds are emitted the room's tuning will be the predominant factor). Accordingly we felt that the most important thing was to leave the artistic aspects of sound to the artists and recording engineers, and on the other hand to specialize speakers so that they would not color such aspects. This we believe will enable optimum reproduction of truly magnificent performances/recordings, and is the concept according to which we implemented the product development.

4 Problems for conventional speakers from the viewpoint of the time domain theory

Here we will consider the issues for conventional speakers when they are looked at from the viewpoint of the time domain theory. First let us remember that for this theory the ideal is for the transfer functions to equal 1, as mentioned above; that is, for the speaker impulse responses to reproduce the impulses just as they are received. Accordingly it is necessary to consider by what causes the waveforms in the impulse responses are distorted and excess signals are added.

4.1 Problems of the speaker box

First we focus on the shape of the speaker. Most conventional speakers have a box shape, which however leads to the existence of fixed reverberation in the speaker because of the generation of standing waves between the facing plane surfaces and because of the natural vibration of the box materials. The standing waves are signals generated later than the original signals; moreover they are not signals that were recorded but rather are signals generated by the speaker itself. Consequently, viewed on the time-base they can be regarded as noise components relative to the input signals.

And the fact that a flat plate is liable to give rise to natural frequencies corresponding to its dimensions, etc., can likewise be regarded as constituting a noise component. Furthermore the flat-plate baffle to which the speaker is mounted distorts the shape of the spherical waves generated from the speaker unit due to the angles in its surfaces. As a result, inaccuracy will occur in the time taken for the waves to reach the listening position. Such factors will cause distortion of the impulse response (refer to Fig. 3).

4.2 Problems concerning mounting of the speaker unit

Most speakers are screwed directly to an enclosure, which results in the speaker unit's vibration being transmitted unchanged to the enclosure via the speaker unit frame, causing fixed reverberation to be generated from the enclosure. And since such reverberation is naturally generated later than the sounds emitted by the speaker, it is another cause of distortion of the impulse response.

4.3 Multi-way unit configuration

Increasing the number of speakers is a highly effective way of making the frequency characteristics as broad and flat as possible, and because of this multi-way speaker units have now become the norm. But though the spherical waves that spread out from each of multiple units in such a configuration should each reach the listening position simultaneously, it is in fact difficult to coordinate them perfectly so that this happens. For instance, with speakers installed so that the tweeter is 20 cm directly above the squawker, if one listens at a position 2 meters away on the axis of the squawker, the squawker's sound will lag approximately 29 microseconds. Converted for, say, a CD sampling frequency of 44.1 kHz, this is equivalent to a lag of 1 sample. This means that relative to each input impulse, the sound will reach the listening point with a lag of 1 sample for each speaker unit frequency band.

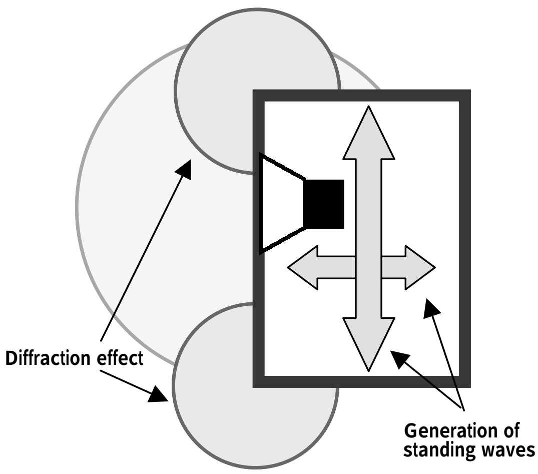

Fig.3 Problem points with the conventional speaker box

Furthermore, with multi-way units, network circuits are necessary for splitting the signals input to each unit into frequency bands, and such circuits will cause distortion of the phase characteristics and consequently of the impulse response. Fig. 4 shows as an example the input signal lag that occurs when a single 3 kHz sinewave is passed through an fc=3 kHz network (secondary LPF/HPF).

This shows the signals input and output to/from the L.P.F. and H.P.F. comprising the network. From this we see that the output waveforms are distorted and that with the same frequency (3 kHz) there are overlapping signals in the space that lag by about 8 samples (fs= 44100 Hz) or 181 microseconds (equivalent to one 5.5 kHz wavelength) after passing through the L.P.F. and H.P.F. This results in these synthetic waveforms having a shape that is greatly altered from the original waveforms (refer to Fig. 5).

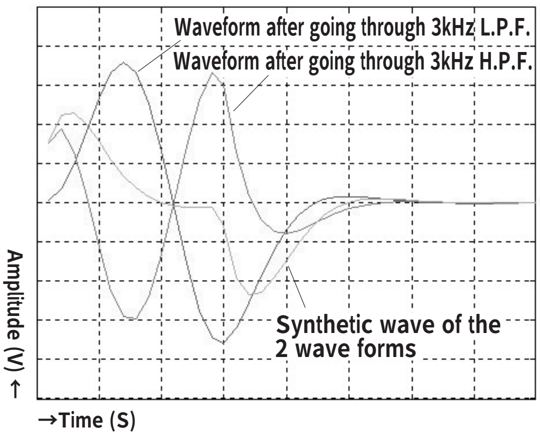

Fig.5 Synthetic waveform of network output

5 Overview of speaker

In order to overcome these problems, our newly-developed speaker has been given the features described below.

5.1 Adoption of a full-range single cone unit

Though single cone speakers are often said by audio fans to produce good sound, in fact with the wider range frequencies now used for audio equipment and music software the fans tend strongly to favor multi-way speakers over single cone, due to the narrow reproduction frequency bands of the latter. But as previously mentioned, single cone speakers are effective at producing good phase characteristics; accordingly the development was oriented toward optimizing the single cone speaker for the unit's diameter / vibrational weight / enclosure volume and taking due account of the structural members, so as to exploit the good phase characteristics while extending the reproduction frequency bands as far as possible.

5.2 Employment of ovoid enclosure

In order to keep to a minimum the standing waves inside the enclosure which are one of the main causes adding excess reverberations to the impulse response, an ovoid shape, which has no plane surfaces in its interior, was used for the enclosure. Ovoid is a shape found in the natural world that has extremely high rigidity, giving it the merit that it is unlikely to generate natural vibration. Furthermore with this shape the baffle surfaces pose almost no obstacle to the spread of spherical waves from the speaker unit, and this enables distortion of such waves to be kept to a minimum.

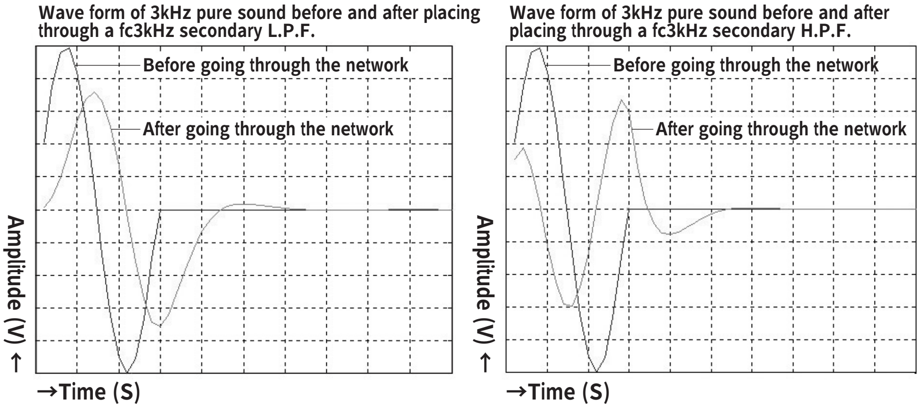

Fig.4 Comparison of network waveforms

Sample data from experiments conducted to confirm the validity of the ovoid enclosure are presented in Figs. 6 and 7. (The speakers used in these experiments were not the newly-developed speaker itself, but items fabricated for experimental purposes.) In the experiments, cuboid and ovoid boxes of identical material were fabricated, identical speaker units were installed to the boxes, and measurements made concerning the resulting sounds. Fig. 6 presents a comparison of the impulse responses and Fig. 7 a comparison of the pulse fall cumulative spectra. From the results for impulse response it can be seen that the cuboid box produces more excess vibration, while the pulse fall cumulative spectra results make it plain that there is continual undesired vibration of around 300 to 700 Hz with the cuboid box. These results confirmed the effectiveness of the ovoid box for suppressing occurrence of undesired sound.

5.3 Mounting of speaker unit on stays

Whereas conventionally speaker units are nearly always mounted by screwing the unit's periphery to an enclosure, for the newly developed speaker a structure is employed whereby mounting is onto stays (refer to Fig. 8).

This accords with the "mechanical ground" way of thinking. In electrical circuitry, the rule is "one point is ground", that is, a specific point is designated as the grounding point. This is a method of suppressing adverse effects such as lowering of the signal-to-noise ratio due to undesired noise flowing into other circuits.

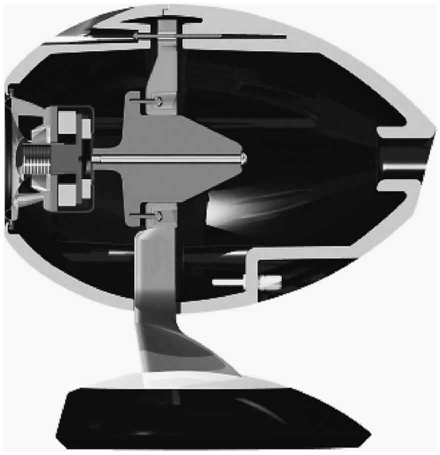

Fig.8 Internal structure of TD 512

The structure utilized in this development constitutes the same approach, with the speaker cone's vibration passed down to the floor via the speaker frame / stays; thus care is taken to prevent generation of excess vibration or sound in the speaker interior. And an "anchor" weighing about 3 kg has been attached to the back of the speaker unit. The idea here is to fix the back of the speaker unit by means of inertial weight, that is to make it into a mechanical ground, thus firmly supporting the back of the unit (the part with the yoke magnet), which is subjected to a reaction when the speaker unit's sound compresses the air, and enabling the speaker cone to compress the air firmly. This has the effect of improving the sensation of speed of the pulse rise and fall, for mid and low sounds in particular.

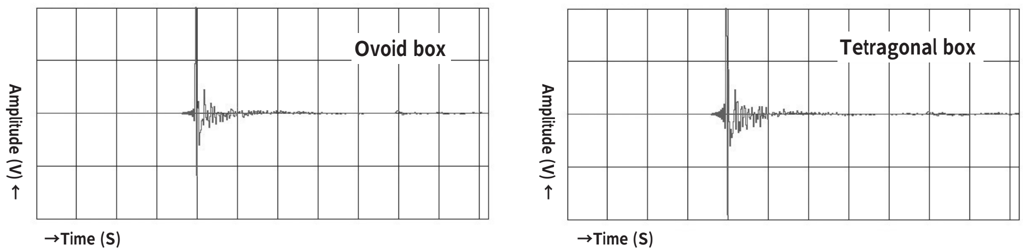

Fig.6 Comparison of impulse responses of ovoid box and tetragonal box

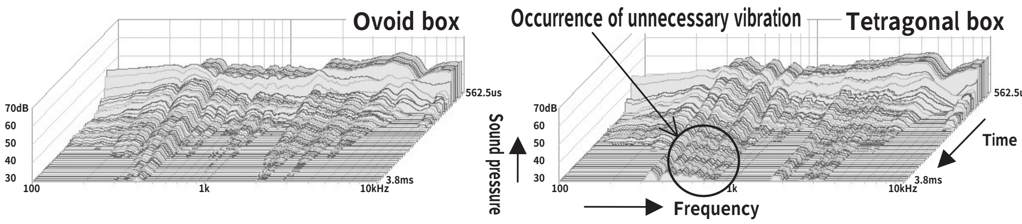

Fig.7 Comparison of cumulative spectra of ovoid box and tetragonal box

5.4 Floating of enclosure

To further minimize the transmission of undesired vibration from the speaker unit into the enclosure, a structure was adopted whereby the enclosure "floats" relative to the speaker unit and stays. In concrete terms this amounted to a policy of interposing a shock absorbent agent between the speaker unit and stays on the one hand and the enclosure on the other, based on the idea of cutting off contact between the two neighboring solids while maintaining airtightness.

Sample data from experiments conducted to confirm the validity of the floating structure are presented in Figs. 9 and 10. (The speakers used in these experiments were not the newly-developed speaker itself, but items fabricated for experimental purposes.) In the experiments, boxes of identical material and shape were fabricated, and into them were fitted identical speaker units, one directly screwed to the box and another installed in a floating condition; measurements were then made concerning the resulting sounds. Fig. 9 presents a comparison of the impulse responses and Fig. 10 a comparison of the pulse fall cumulative spectra. From the results for impulse response it can be seen that the floating structure produces less excess vibration, while the cumulative spectra results make it plain that there is continual undesired vibration of around 300 to 500 Hz with the directly-mounted type. These results confirmed the effectiveness of the floating structure for suppressing occurrence of undesired sound.

The major specifications of the newly-developed speaker incorporating the technology described above are as follows:

Reproduction frequency characteristics (-10dB): 40-17 kHz
Impedance: 6Ω
Sound pressure level: 81.5 dB / W・m
Allowable input (rated / maximum): 30W / 60W
Outer dimensions (mm): W286×H372×D364
Weight: 14.2 kg

6 Overview of the amplifier

Concurrently with development of the new system, a new amplifier was developed in order to exploit the speaker's capability to the full. An overview of this is presented below.

6.1 Separate power unit and amplifier unit

This amplifier is divided into a conical amplifier unit and a cylindrical power unit, and is designed so that the amplifier unit can be mounted on top of the power unit.

Amplifiers of this class generally have the power unit and amplifier unit integrated together, which however makes undesired vibration generated by the transformer itself liable to be transmitted to the amplifier circuits, superimposing undesired vibration on the audio signals and thereby causing a deterioration of the sound quality. Therefore we opted for the separated configuration.

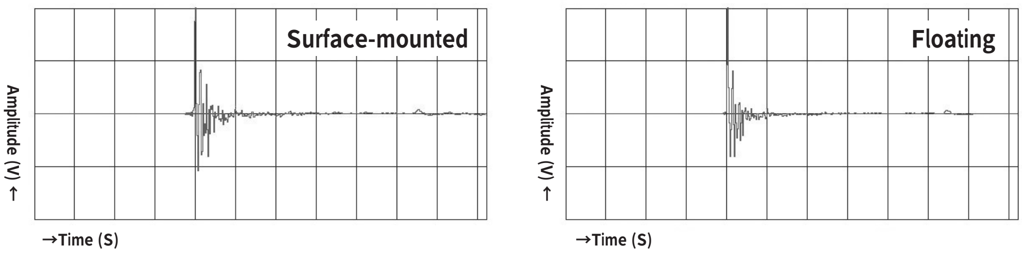

Fig.9 Comparison of impulse responses of surface-mounted and floating types

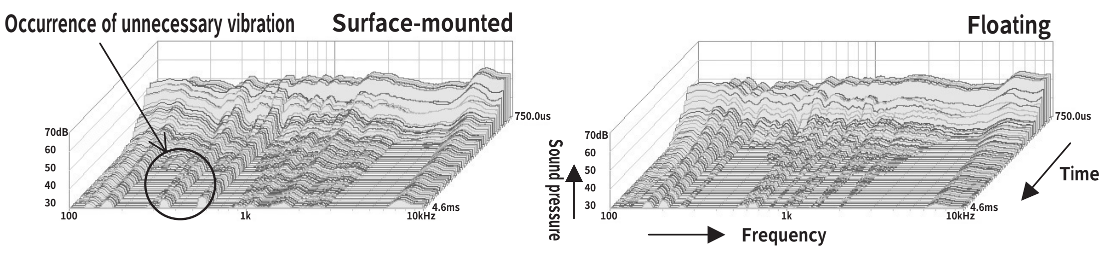

Fig.10 Comparison of cumulative spectra of surface-mounted and floating types

6.2 Employment of aluminum body

This product was developed with emphasis on eliminating vibration. Although it is normal to use steel plate for the body, we elected to use aluminum material instead for the following reasons:

(1) There are limits on the thickness of steel plate that can be used, and there is risk that the steel plate itself will vibrate.

(2) Steel plate is lacking in formability, and thus allows little freedom in design.

(3) When steel plate is used, the power ICs, which are a heat emission source, are screwed directly to the amplifier unit body and thus the body also serves as a heat dissipater.

Further, as mentioned above a conical shape is used for the amplifier unit, thus heightening the rigidity of the body. And it has contributed to the creation of an original form.

6.3 Simplification of circuits

Simplification was also implemented for the internal circuits, in accordance with the time domain theory which holds that the recorded sound information should be reproduced as faithfully as possible, without any coloring. Tone control and equalizers such as found in other companies' amplifiers were totally eliminated, with the development focusing solely on simple amplification of sound. Moreover there is just a single input system and control is by dials only; thus functions too have been pared down to a bare minimum. This enables unadorned, extremely straightforward sound reproduction. Hence this amplifier can be said to be optimal for sound reproduction by this system's speakers.

The major specifications of the newly-developed amplifier incorporating the technology described above are as follows:

Rated output: 30W (when T.H.D. = 1%)×2 channels
Input impedance: 10 kΩ
Load impedance: 6Ω or more
Reproduction frequency response (-3dB): 10-100 kHz
Higher harmonic distortion rate: No more than 0.05%
Outer dimensions (mm): Main body W286×H372×D364
Power unit W215×H97×D187
Weight: Main body 2.6 kg
Power unit 3.3 kg

7 Acoustic characteristics

The acoustic characteristics of this speaker are as described below.

7.1 Impulse response

The impulse response of the newly-developed TD-512 is shown in Fig. 11. Features are the facts that there is almost no pre-echo compared to other speakers, and that overshoot is low. Thus pulses for almost all frequencies rise and fall in phase.

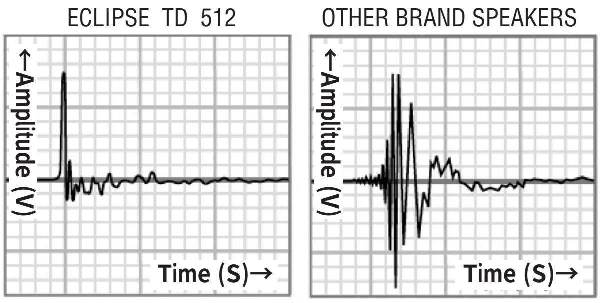

Fig.11 Comparison of impulse responses

7.2 Sound pressure frequency response

The sound pressure frequency characteristics of the newly-developed TD-512 are shown in Fig. 12. Since the impulse response is neat as the above data show, naturally the sound pressure frequency characteristics are flat.

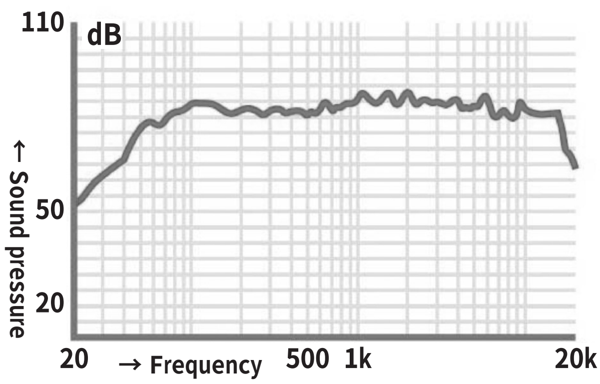

Fig.12 Frequency response of the TD 512

7.3 Pulse fall cumulative spectra

A comparison of the pulse fall cumulative spectrum of the newly-developed TD 512 with such spectra of other companies' speakers (2-way cuboid box type) is given in Fig. 13.

8 Conclusion and discussion

In the present product research we took shape to be an important transfer function, and in order to make the transfer functions approximate as closely as possible to 1 we made innovations not only to the speaker unit shape but also to the enclosure's shape, as well as to the support structure. In this way we endeavored to create a new type of sound.

As a result the pulse rise and fall of the sound was made perceptually quicker, thus enabling clearer reproduction of exquisite and emotional performances of professional artists, including for example sounds produced by sensitive keystroke piano playing, the delicate play of bow on strings, or mouth movements in vocal renderings. And reproduction of the sound field has been made sharp and real not only in the lateral direction but also in the depth and vertical directions. This is because whereas with conventional speakers impulse response containing multiple pulses generates multiple sound waves resulting in multiple sound images, the new system has the feature that when there is only a single pulse, only a single sound image results (refer to Fig. 14).

The obverse of this however is that depending on the recording conditions, differences will occur in the realism of the sound field. For example with a recording of a live performance by acoustic instrument the recorded signals will all be of natural sounds and accordingly the volume of information will be large, so that the details will be reproduced with high realism, including the position information; but with music that has undergone heavy acoustic editing such as that involving electronic instruments or equalizing, the sound field reproduced will be lacking in 3-D feel due to the processing of the natural sound information, which will also create a tendency to coarseness in the recording/editing. Put simply, this is due to the fact that the transfer factors approach 1, which means that all the recorded information is reproduced without passing through any filter. To give a metaphor for this phenomenon in terms of images: a lens with a shaper degree of resolution will make visible things that were hitherto unseen.

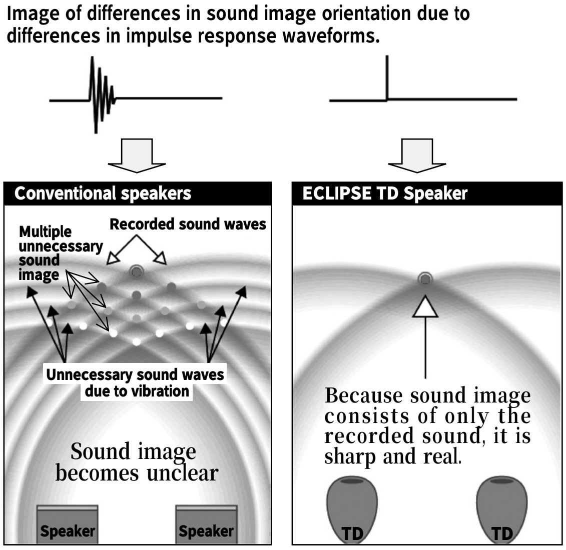

Fig.14 Schematic of sound image orientations resulting from differences in impulse response

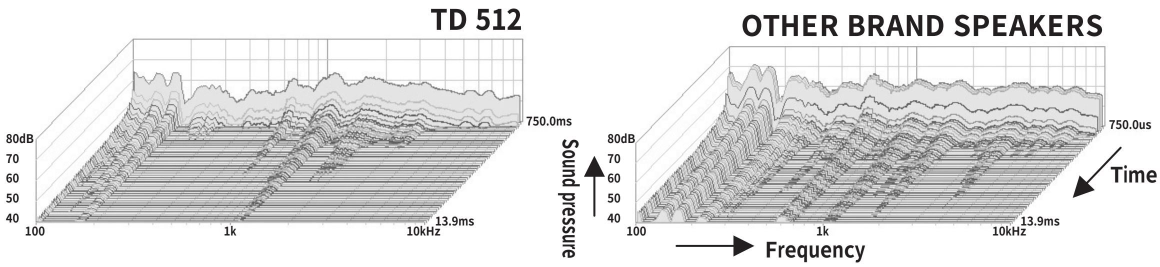

Fig.13 Comparison of cumulative spectra

In future research we intend to proceed with investigation for further quantifying the physical characteristics of the speaker, to make further improvements to the time-base characteristics, and to extend the reproduction frequency bands.

Finally we would like to thank President Hiroyuki Yoshii of Time Domain, and all the others at that company who gave us guidance in the development of this speaker, and to express our gratitude to all at Foster Electric CO.,LTD. for their assistance in the development.

Profiles of Writers

Akira Nishikawa

Joined Fujitsu TEN in 1990. Involved in the development of car sound systems. Currently part of the AE Project of the Acoustic Engineering Department at the Audio Business Division, A.V.C. Products Group.

Isao Yokoyama

Joined Fujitsu TEN in 1991. Involved in the development of car audio equipment. Currently in the Home Audio Business Department of the Audio Business Division at A.V.C. Products Group.

Katsuhiko Tsumori

Joined Fujitsu TEN in 1986. Involved in the development of car sound systems. Currently part of the SP Project of the Acoustic Engineering Department at the Audio Business Division, A.V.C. Products Group.

Hiroshi Kowaki

Joined Fujitsu TEN in 1985. Involved in the development of car engine control systems, and the develoment of basic audio technology, such as digital signal processing. Currently the project Manager of the Home Audio Business Department at the Audio Business Division, A.V.C. Products Group.