DIGITAL AUDIO EQUIPMENT FOR PROFESSIONAL USE

Toshiyuki Takegahara

NHK Science and Technical Research Laboratories, Tokyo, Japan

ABSTRACT

In quest of higher quality, stability, smaller size and lower cost, we have advanced from the age of vacuum tubes to that of transistors and then to that of ICs. And such technological progress has made it possible for us to make extensive use of digital technology in recording and in broadcast engineering.

At present, in the 'package' sector, PCM recorders are used and digital disks are providing us with programs of high quality. In transmission systems, program circuits have been digitized, with the result that programs with high-quality sound can now be enjoyed on FM broadcasts throughout Japan. TV broadcasting via satellite has already been put into practice, in which the PCM system is adopted for sound. With this in mind, we would now like to consider the present condition and future outlook of digital equipment for professional use, focusing our attention on the complex broadcast technology in this system and also touching upon questions of recording techniques.

1. CHARACTERISTICS OF DIGITAL EQUIPMENT USED IN THE ANALOG SYSTEM

One of the objectives of introducing digital equipment into the analog system is to achieve better quality. Fig. 1 shows the dynamic range of each item of equipment constituting the audio system for program production. From Fig. 1, it is noted that the analog audio tape recorder and the analog disk player represent the two weak points in the system as a whole. This means that, at present, introduction of a PCM recorder or a digital audio disk player would be most effective in achieving higher sound quality. For that reason, since 1971 the PCM recorder has come to be used for original recording, and LP (long playing) records known as PCM records have been produced. In broadcasting, too, PCM records have been used in FM broadcasting services since 1977.

The second advantage of digitization is found in high stability. Digital announcing machines are used not only in broadcasting but in other services as well. Those machines are used for transmissions of emergency warning commentaries, as well as call signs and notices. Even though they are not used as frequently, high reliability is expected of them because of the need for such messages to be transmitted without fail. In the past, a recording/playback system with a tape cartridge has been used, but a high-reliability device, without a mechanical section and using solid-state memory such as ICs or magnetic bubbles, was brought into practical use, even though the sound quality was just good enough to enable the listeners to make an accurate judgment of the information produced by a digital compression method such as delta modulation (ΔM) and differential PCM (DPCM).

Another advantage of digitization is enhancement of the functions of the equipment. There are, for example, such devices as audio files and CM banks, both of which are a combination of the data-bank portion with many other functions such as retrieval, editing and transmission. Some of these devices are so huge that they occupy a whole room. Up to the moment when a program is actually sent out, pieces of information are accumulated in the device and it takes a considerable time to retrieve the necessary portions and to edit them into a program. However, as a result of digitization, the sound data, even after compression, can be kept in a medium-level quality and, by storing the data in a large fixed disk, it is possible to complete the retrieval and editing and send the program out at a speed incomparably faster than the conventional method.

One typical item of equipment that has been digitized to achieve high performance, multi-function, and reduction in size, all at one time, is the reverberator. In the analog system, there are such devices as the echo chamber, the spring and steel-sheet reverberator. However, an echo chamber is of course immovable and the spring-type or steel-sheet-type reverberator, though transportable, is not only unwieldy but also vulnerable to exterior vibrations and noises. In contrast to these, the digital reverberator has been greatly reduced in size and can be transported easily. It is also possible to set the timing of reflected sound and the parameters, such as the attenuation characteristics, with ease. Using this digital device, it has now become possible to produce special reverberation effects that actually do not exist in the physical world as well as more natural reverberation.

2. THE NECESSITY OF DIGITIZING THE ENTIRE SYSTEM

Thanks to the developments made in PCM broadcasts by satellite and in digital audio disk technology, we can now easily enjoy radio programs of high quality in our homes. The audio system on the production side of these programs is almost entirely analog with the exception of the PCM recorder. Still, its performance is well above the level of performance obtained by 16-bit quantization and, in fact, A/D and D/A converters have become the factors that determine the characteristics of the entire system.

In the case of mixdown, for example, deterioration in sound quality is caused by the increase in quantizing noise resulting from the repetition of A/D and D/A conversions, as well as by high-pass group delay distortion resulting from the repeated passage of the sound through the low-pass filters used in A/D and D/A conversions. The digital equipment used in homes today to receive the final programs is also of 16-bit performance, the same as the equipment for professional use. Hence, there is need to pay closer attention to deterioration in sound quality at the production stage than in the case of an analog system. Therefore, it is desirable that the signals, once digitized, should be processed as digital signals up to the final reproduction stage.

3. COMPOSITION OF A DIGITIZED SYSTEM

Fig. 2 shows a conceptual chart of a basic audio system for professional use. All items of equipment constituting the system are interlocked with the master clock. The signals unrelated to the master clock, such as the signals with a different sampling frequency, are synchronized with the master clock by converting the sampling frequency, while the signals generated at a remote place, such as those on junction line, are given phase modification to match the phase of the master clock by the use of a synchronizer.

The most important element in the system is the master clock for which the frequency is generally set at an integral multiple of the sampling frequency. Also, the interlocking of this sampling frequency with the clock frequency used for video is indispensable in maintaining video-audio synchronization. Fig. 3 shows an example of an interlocking device for professional-use sampling-frequency and TV synchronization.

In some of the bigger systems, the digital signals generated by different master clocks are synchronized, even if the nominal value of the sampling frequency may be the same. In such a case the signal processing should preferably be done by adjusting the phase with the clock that is playing the central role in the system. It is for such a purpose that the digital synchronizer for audio is utilized. In adjusting the phases of two clocks, what is normally done is either to subtract or add to the sample the difference with the sampling frequency. In this case, the more desirable method would be the one that causes less change in the pitch of the sound of the program or less time-lag in the performance hour and that which maintains the audio-video lip sync. An interpolating-type digital synchronizer using a digital filter and a delay-switch type are currently being proposed to serve the purpose as mentioned above.

In the case of a digital system for professional use in a studio, the specified sampling frequency is 48 kHz, whereas the frequency of circuits and Compact Discs (CD) are 32 kHz and 44.1 kHz, respectively, with the result that conversion of the sampling frequency will be required. Different conversions, even one for a 48 kHz/32 kHz conversion at a small converter ratio of 3:2, will inevitably result in degradation in sound quality by several dB in the dynamic range. In the case of a broadcasting station where a complex system is used, a device that causes such degradation in characteristics as mentioned above should preferably be inserted as much as possible at the last stage of the system.

The most appropriate method of determining the recording level of digital signals is to set it at the maximum value of the digitized codes without exceeding the maximum level. As a result, it sometimes happens that audibility level varies from program to program.

In the case of broadcasting in which all kinds of programs such as music, news and dramas are transmitted one after the other, it is essential that the audibility levels of all the programs are kept basically the same, otherwise listeners would be obliged to adjust the volume of their receivers each time a different program is received. For that reason, in fact, it has become impossible to maintain compatibility in the audibility level between a digital-disk program and the PCM broadcast.

One of the possible solutions to this problem would be to adopt a method in which the level-control signals are put onto the digital disk so as to make it at the same level as that of the broadcast, and to control the output of the analog signals of the player at the time of playing the digital disk. But to insure stable service at a location at the fringe of the service area of the satellite, we would recommend the method of inserting in the final stage of the transmission system a digital device which worked in the same way as the compressor does in the AM broadcasting.

As shown in Fig. 4, fixing the audibility level as it is perceived by the human ear is attempted by selecting the curve according to the program being listened to. Action to extend the level is done in portions where the level is low and then to compress the level in portions where the level is large. Also, depending on which curve is being used, it becomes possible at the receiving side to easily revert the characteristic to linear by multiplexing on the audio data.

4. FUTURE OF THE DIGITAL AUDIO SYSTEM FOR PROFESSIONAL USE

The introduction of digital technology has facilitated enhancing performance of equipment and maintaining and insuring characteristics of each item of equipment. In the future, audio systems in their entirety will be digitized and, moreover, will become equipped with auto-check functions. Furthermore, the advent of new systems using digital technology and new digital equipment may be expected.

One of the fields in which the practical application of new digital systems is expected is Hi-vision (high definition television) whose selling points are high-quality pictures and sounds. As a result of the development of satellite broadcasting and video disks, it is going to become possible for people to enjoy in their homes not only the video pictures of high resolution on a large display but also multichannel sounds by PCM. In the future, Hi-vision will be shown in theaters on wide screens similar to those used at movie houses.

As for an audio reproduction system, the 4-channel reproduction system, which offers the sensation of being in a concert hall or the sound effects like that of cinerama, is quite promising. Since this 4-channel reproduction system, unlike 4-channel stereo for phonograph records, has naturally been digitized, multiplexing sound-field control signals, which used to be difficult with analog technology, can be done easily. This means that the reproduction of sounds with this 4-channel system may be done freely varying the normal sounds and by the feeling of spreading the sound according to the kind of program and to the picture on the screen.

In the case of a large-scale audio system such as those in a theater, many loudspeakers can be set up around the audience seats and, by multiplexing the signals that select the sounds coming from four of those loudspeakers on the picture frames, the audience would be able to enjoy the surrounding effects of sounds together with the changing pictures on the TV screen just as they do at a theater showing cinerama.

In the case of a smaller-scale system in the private home, it is possible to convert the signals into 2-channel or 4-channel signals based on the sound-field control signals, which are multiplex transmitted and which produce a kind of sound effect never dreamed of with conventional stereophony. It is also possible to easily make a system in which many loudspeakers are set up. It is, indeed, digital technology alone that enables us to create a method capable of dealing with a system of any scale, from the largest to the smallest.

One of the advantages derived from digitization of sound signals is that multiplexing programs can be done easily. Studies are being made at present to conduct a multichannel PCM sound broadcasting using one satellite channel. On such a PCM broadcasting, a large number of programs with medium-level sound quality, such as information on living including news and weather forecasts, in addition to the music programs that require high-quality sound, can be multiplexed.

In the case of a digital audio system, characteristics can be freely selected and designed by means of selection of quantization numbers and sampling frequencies. In the case of a transmission system such as a satellite in which the transmission capacity is restricted, it is extremely important to make an appropriate allotment of quantization numbers and sampling frequencies in securing sound quality that is superb in terms of audibility.

The sampling frequency determines the transmission bandwidth of sound. The equi-quality curve shown in Fig. 5 represents the results of the studies made on how sound quality changes according to combinations of sound transmission bandwidth and the quantization numbers. Using this curve, the optimum quantization number and transmission bandwidth in a digital system can be obtained.

In order to obtain the desirable quantization number and transmission bandwidth for a digital transmission system by using this equi-quality curve, it is appropriate to use an equi-quality curve of the type of sound on which strict evaluation is normally done, such as that of the shamisen (a Japanese musical instrument like a banjo).

If the volume of information transmitted in a digital transmission system is expressed by Q kbit/s, the transmission bandwidth of B and the quantization number of N, then the following relation will be established: $N\ B = Q/(2+\alpha)$. Here, α is a constant that is determined by the performance of the low-pass filter used in the digital circuit.

As an example, the relationship between the quantization number and the transmission bandwidth that ensures maximum quality in the case when Q=180 kbit/s and $\alpha = 0.2$ are shown as curve A on the equi-quality curve.

While the quality in terms of audibility changes in the range from 1 to 3 in evaluation, the maximum evaluation of 3 is obtained at point P. At that point, the quantization number 11.8 bit and transmission bandwidth 7 kHz can be read from the chart. Therefore, to obtain maximum quality, the quantization number should be set at 12 bit and the sampling frequency, 15 kHz. In the future, when the sound programs come to broadcast using a multiplex system, the sound quality will be determined by the selection of the quantization number and sampling frequency. So, it is important to pay careful attention to this selection.

Furthermore, digital technology has made complex processing of signals possible, the kind of processing that analog technology was unable to handle. So, from now on, expectations can be placed on the advent of a new type of effects machine using the type of technology as mentioned above.

The design of conventional effects machines has been based on filtering and delaying. The digital effects machine which has now been put into practical use has been developed along the same line.

For example, the voice in which the low sound is emphasized and the high-frequency component is weak lacks clarity, and becomes a voice with poor audibility. In such a case, the clarity of the voice can be improved by inserting a low-cut filter, suppressing the low-frequency component and raising the high-frequency component with an equalizer. However, when it comes to improving the audibility of a voice that lacks intonation and clarity, the conventional method is totally ineffective. But even such a type of a voice can be improved with digital signal processing. Let us introduce an example.

If we were to give a typical example of a clearly audible voice, we would say that it is an announcer's voice. Fig. 6 shows a comparison of an announcer's voice with that of an ordinary one. In the case of the ordinary voice, the pitch interval is the same in portions (1), (2) and (3). In the case of an announcer, the waveform of the intervals at (1) and (3) is more sparse than at (2). This means that the intonation of an announcer's voice is bigger than that of an ordinary voice.

If an effects machine capable of changing the pitch frequency of a voice could be built, it would be feasible to increase the intonation of an ordinary voice and process it into a voice with its audibility improved to some extent.

The human voice is produced by the vibration of the vocal cords. The vibrated air, as it passes through the vocal resonance path, is filtered and is issued from the mouth as a voice.

In changing the pitch frequency of a voice, we first must produce the signals consisting only of the pitch-frequency of the voice from which the influence of the filtering effect of the vocal resonance path is eliminated. The processing done here is extracting the residual signals that are the pulse signals with the same cycle as that of the pitch of the voice. This can be done by dividing the voice within a time period short enough to keep the characteristics of the vocal resonance path from changing substantially and by putting the voice through a filter that has the characteristics opposite to those of the vocal resonance path. Fig. 7 shows this process in terms of analog signals.

In order to change the pitch frequency, the residual signals should be further divided by pitch cycle and the pitch cycle should be reformulated by adding dummy signals to each section or eliminating the tail portions. As shown in Fig. 8, the pitch information thus processed is combined with the speech signal after putting the signals through a filter with characteristics opposite to those of the filter used at the time analysis, that is, the filter with the same characteristics as those of the vocal resonance path.

Since the characteristics of the vocal resonance path change constantly according to the changes in the shape of the opening of the mouth, it is necessary to make adaptations every moment. This kind of processing is possible because the voice can be digitized and computer-processed, and a device which is capable of this may be regarded as a new type of effects machine.

Fig. 9 shows an example of the voice reformulated through the abovementioned process. (A) represents the original voice and (b) the voice after its pitch variations have been vastly reformulated. As a result of processing the signals as explained above, the voice has been reformulated into one with modulated intonation.

Although real-time processing is difficult at the present stage, a new type of effects machine capable of improving the clarity of a voice may become a reality if the speed of signal processing could be further increased by either reducing or removing the reverberation component from an unclear voice accompanied by reverberation. Until such equipment is a reality, various types of processing would be required, such as, splicing the divided signals. However, this could result in such problems as an interruption at the spliced portions or errors in the estimation of the vocal resonance pass characteristics. So, finding solutions to these problems would be a task to be undertaken in producing high-quality sound, along with the effort to develop equipment capable of real-time processing.

5. AFTERWORD

It was in the tape recorder in the package sector that digitization was achieved first, and this was followed by digitization of the transmission system which had been troubled by exterior noises. The improvements seen in the performance of digitized systems are acknowledged by everyone.

In order to enhance the performance of entire systems, it is of course necessary to digitize the systems completely. It would, however, be totally meaningless if the merits innately possessed by the digital technology, such as high stability, high functions, compactness, and low price, could not be brought forth. It is not enough merely to translate, as has been done heretofore, analog equipment into digital equipment which, even in its present state, has performance of 16 bits or higher.

What needs to be done is to devise a digital system that is compatible with an analog system, by making effective use of the merits of analog and taking into full consideration the technological advantages that only digital can achieve. For that purpose, the author hopes that the wisdom of the engineers of the world will be brought together and, based on the accurate appraisal of the status quo and the future outlook, establish uniform standards and specifications on a worldwide basis, so that all the countries of the world may become able to freely exchange broadcast programs.

BIBLIOGRAPHY

*Signal Synchronization in Digital Audio, W. T. Shelton (The 76th AES)

*A New Approach to Sampling Rate Synchronization, Roger Lagadec (The 76th AES)

*Trial Manufacture of a Digital Audio Synchronizer Using Digital Interpolating Filter, Watanabe and Takegahara (March 1985, Japan Society for Acoustics, 2-3-15)

*Problems of Digital Audio Systems in Broadcasting, Tanabe, Watanabe and Suganami (Feb. 1985, Japan Society for Electronic Communication, EA84-75)

*PCM Sound Broadcasting Systems via DBS, Yoshino ('85 Tokyo AES)

*Influence of Group Delay Distortion of Low-Pass Filters on Tone Quality for Digital Audio Systems, Hoshino ('85 Tokyo AES)

*Optimum Condition of the Quantizing Number of Bits and the Sampling Frequency for a Digital Transmission System with a Constant Bit Rate, Takegahara (The 75th AES)

*Analysis of Announcer's Voices and the Effects of Training, Kuwabara and Ohgushi (Oct. 1984, Japanese Society for Acoustics)

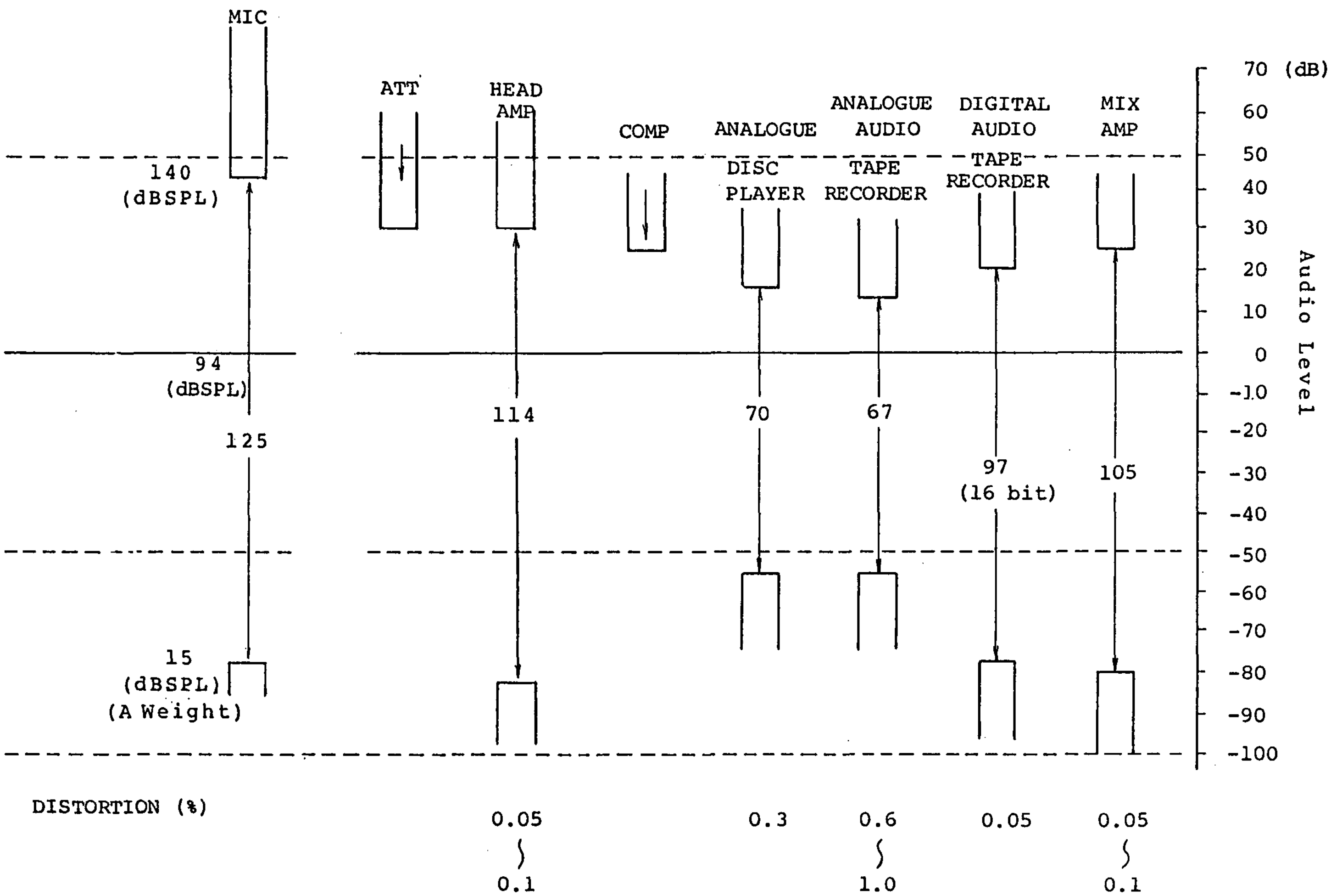


Fig. 1 Dynamic range of audio equipment.

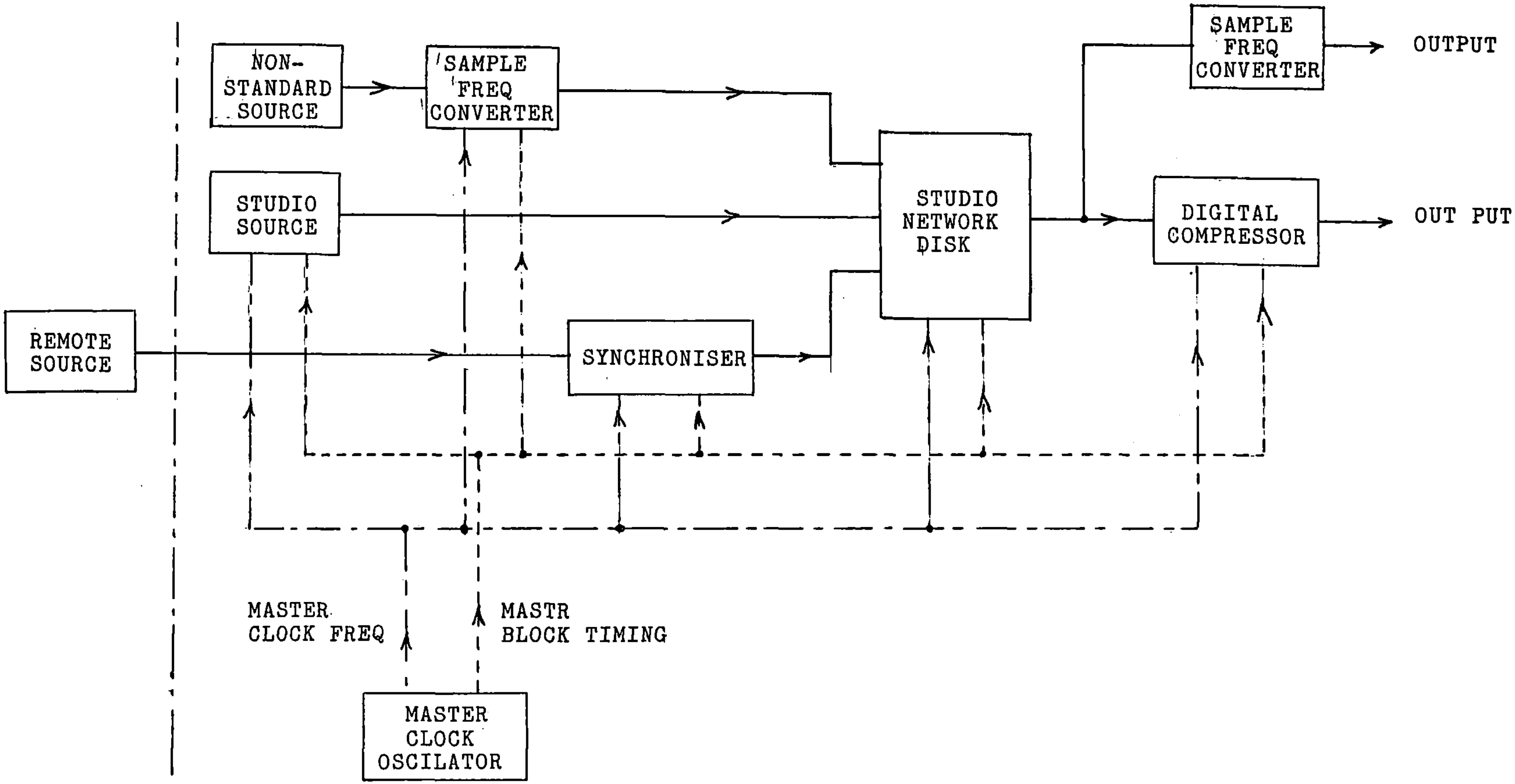
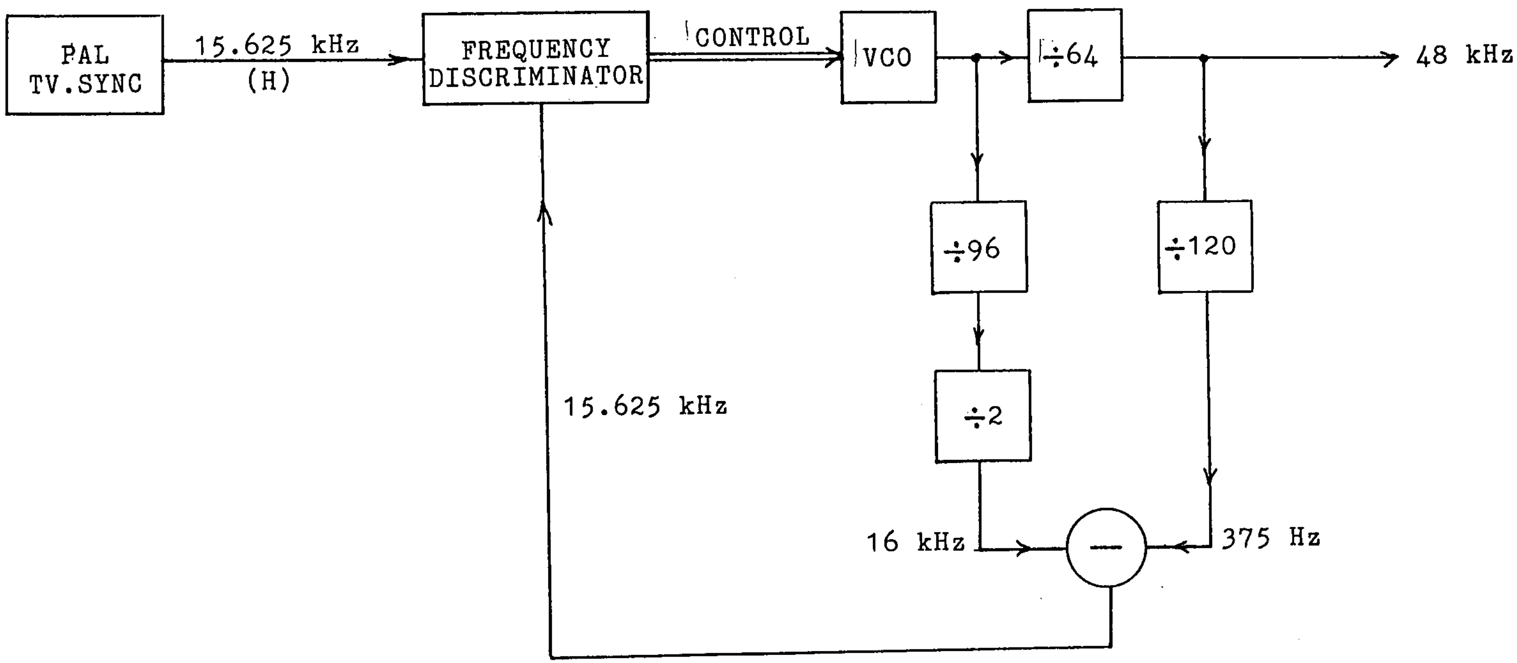
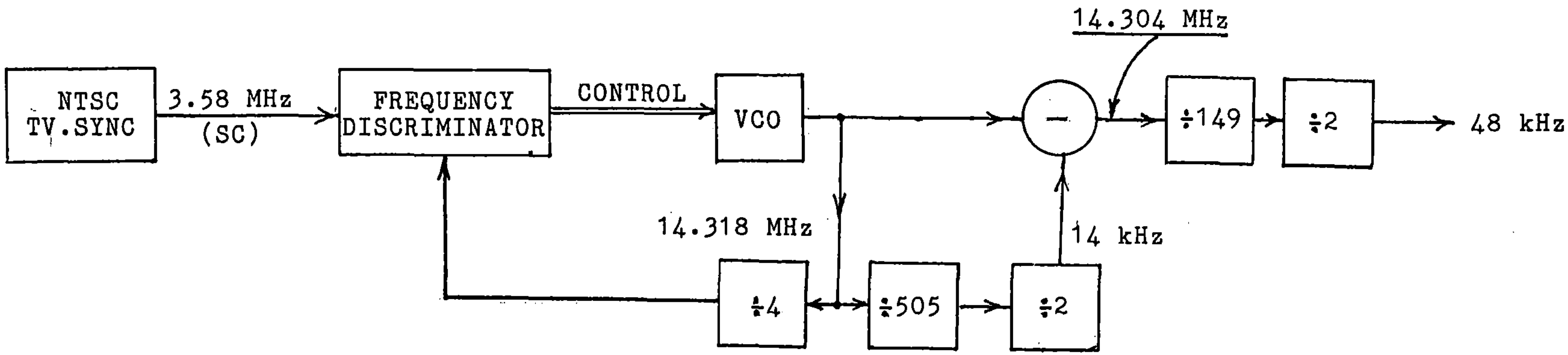


Fig. 2 Digital audio studio system.



a) For PAL system.



b) For NTSC system

Fig. 3 Sampling frequency oscillator for locking TV sync

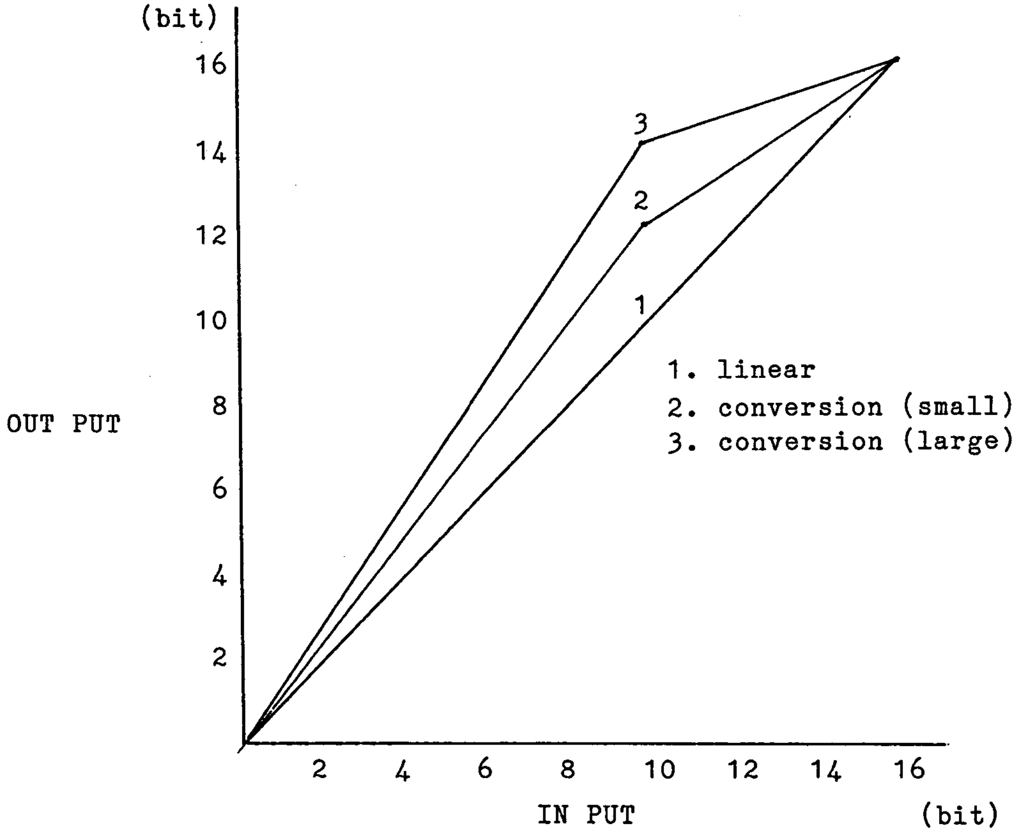


Fig. 4 Bit conversion characteristics.

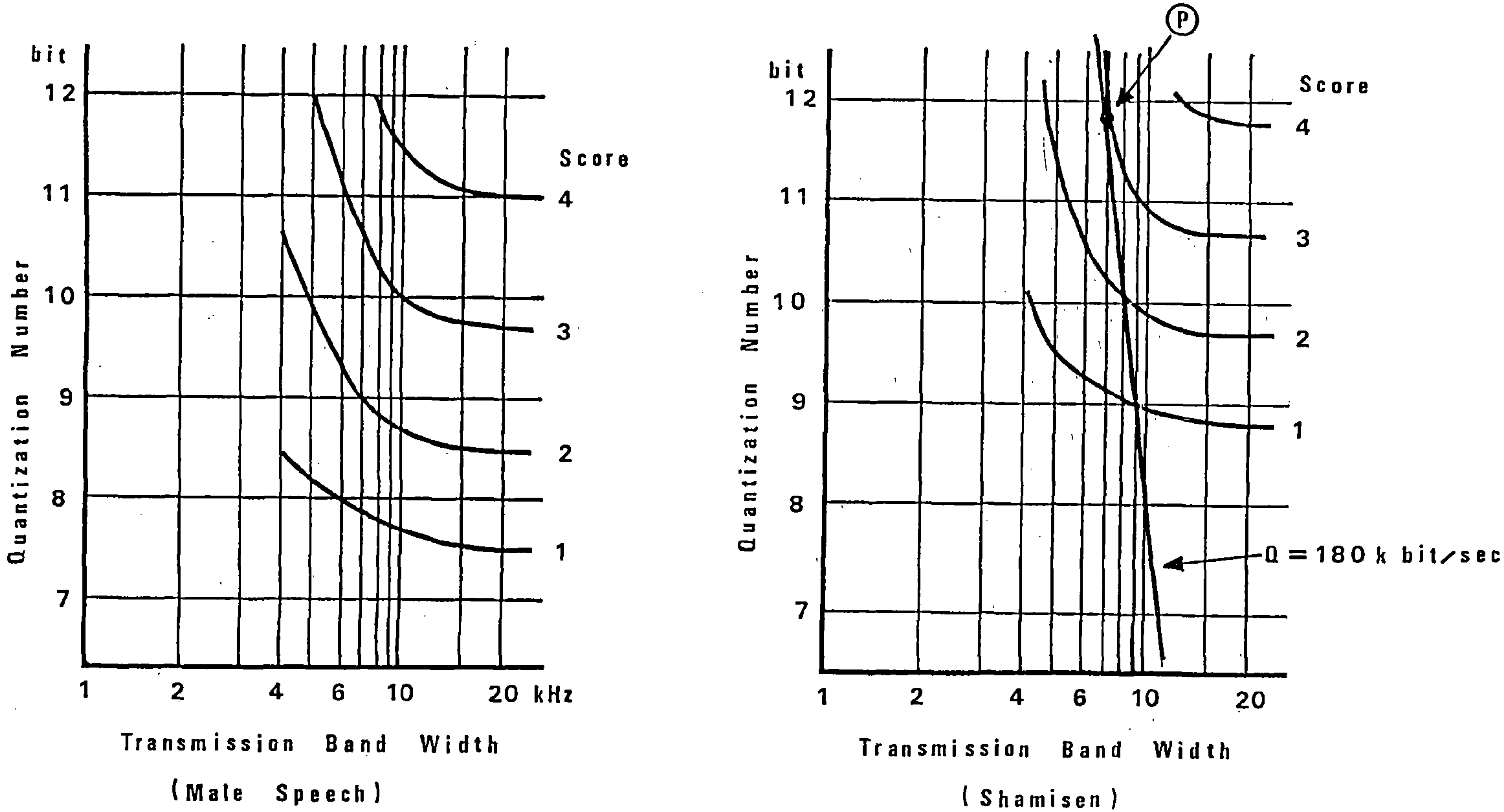
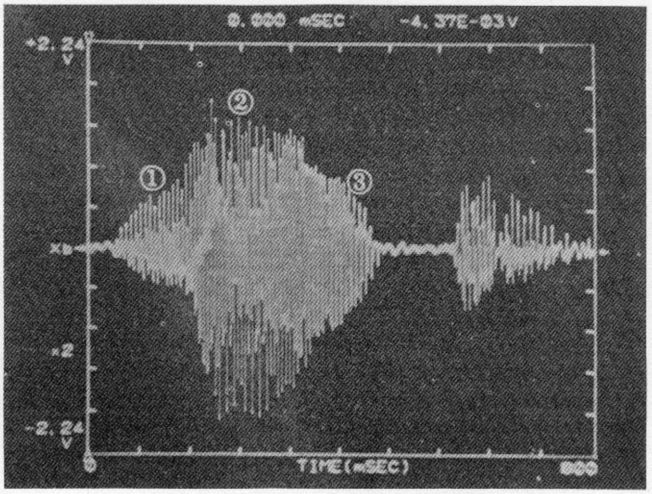
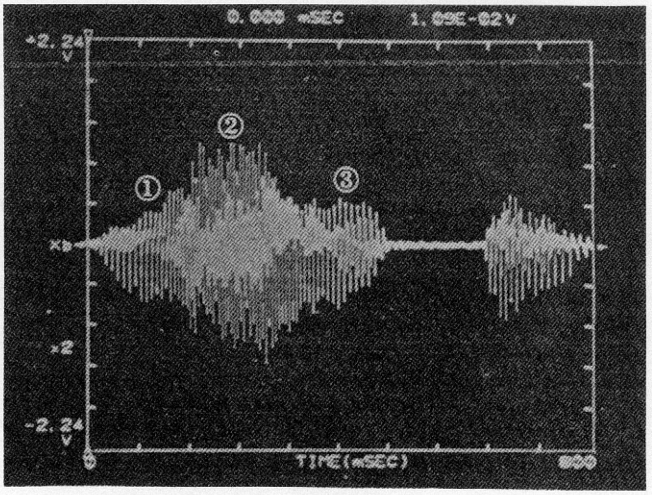


Fig. 5 Equi-quality curve and 180 kbit/sec transmission rate curve.



generality



announcer

Fig. 6 Waveform of voice

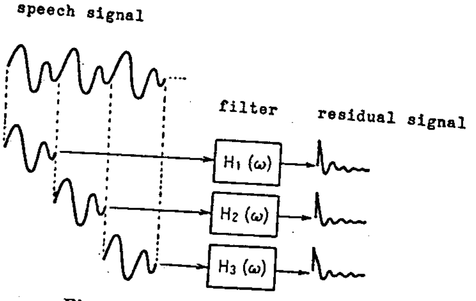


Fig. 7 Pitch frequency analysis.

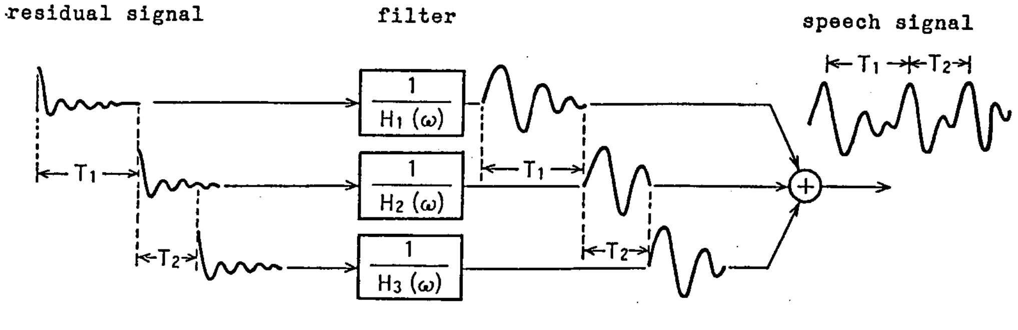
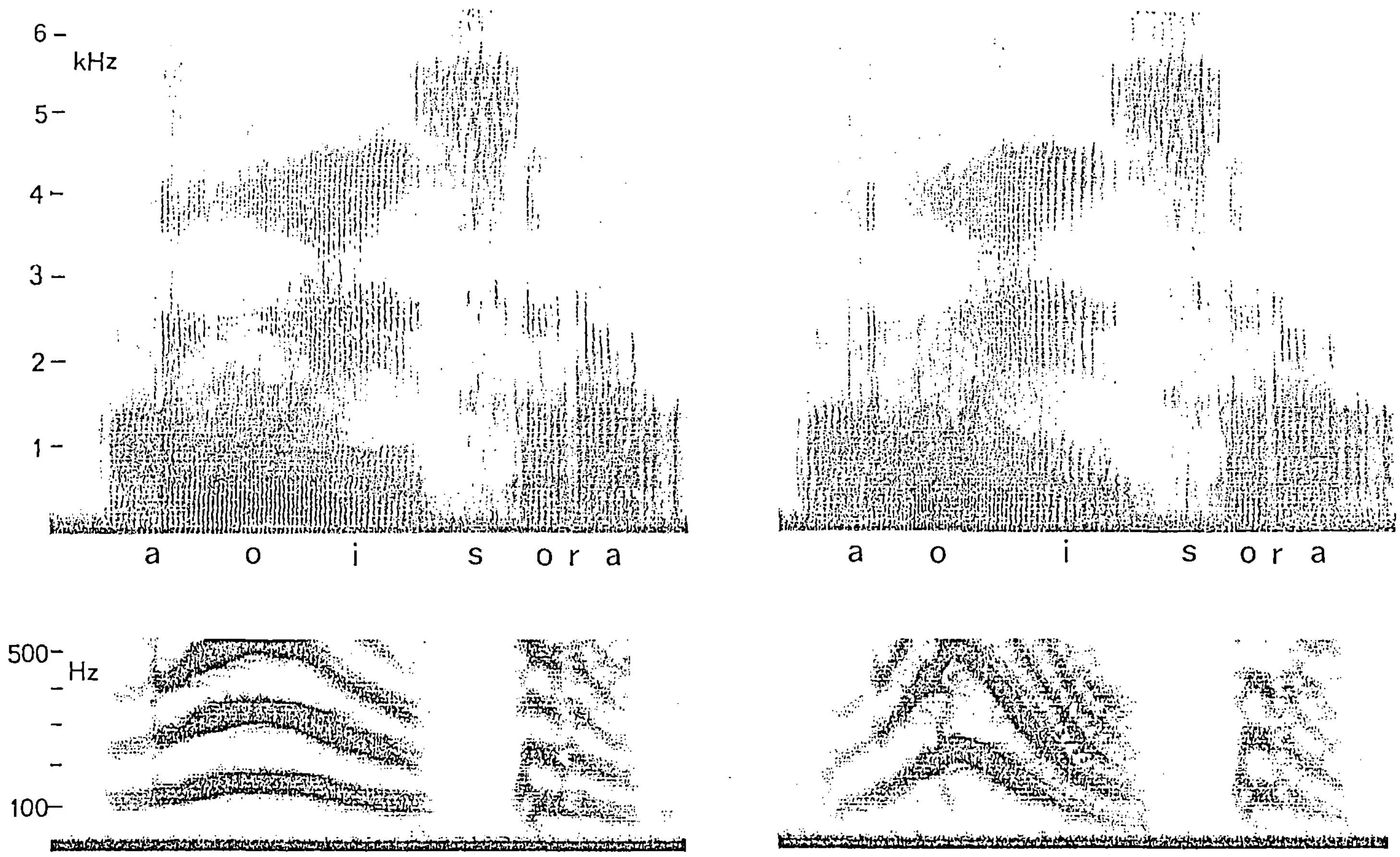


Fig. 8 Pitch control synthesis.



(a) (b)

Fig. 9 Sona-graph (uper) and pitch pattern (lower).