

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 February 2003 (27.02.2003)

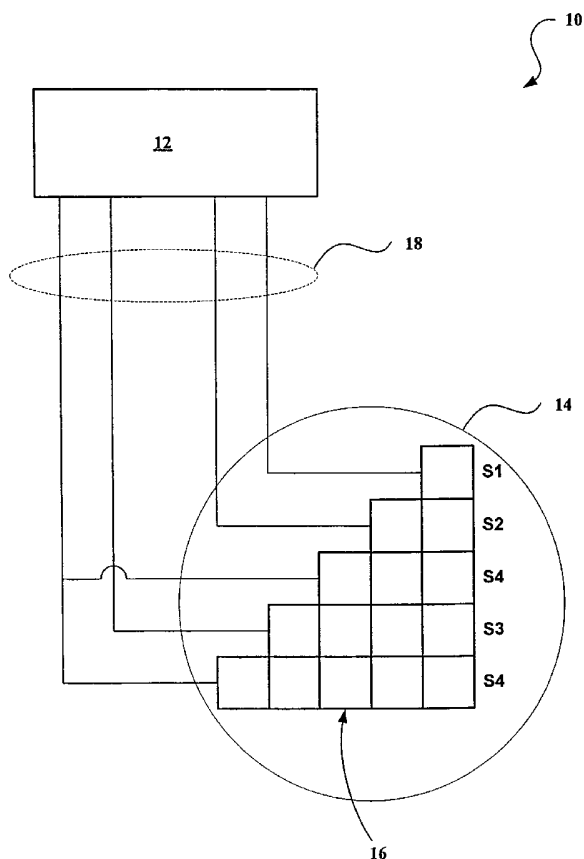
PCT

(10) International Publication Number
WO 03/017717 A2

- (51) International Patent Classification⁷: **H04R**
- (21) International Application Number: PCT/US02/26290
- (22) International Filing Date: 16 August 2002 (16.08.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/313,379 17 August 2001 (17.08.2001) US
- (71) Applicant: **CARNEGIE MELLON UNIVERSITY**
[US/US]; 4615 Forbes Avenue, Pittsburgh, PA 15213 (US).
- (72) Inventors: **KAIGHAM, Gabriel, J.**; 5501 Aylesboro Avenue, Pittsburgh, PA 15217 (US). **NEUMANN, John, J., Jr.**; 548 Loretto Road, Pittsburgh, PA 15217 (US). **DIAMOND, Brett, M.**; 20 Throop Avenue, Monroe Township, NJ 08831 (US).
- (74) Agents: **PENCOSKE, Edward** et al.; Thorp Reed & Armstrong, LLP, One Oxford Centre, 301 Grant Street, Pittsburgh, PA 15219-1425 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR RECONSTRUCTION OF SOUNDWAVES FROM DIGITAL SIGNALS



(57) Abstract: Each of a plurality of speaklets (MEMS membranes) produces a stream of clicks (discrete pulses of acoustic energy) that are summed to generate the desired soundwave. The speaklets are selected to be energized based on the value of a digital signal. The greater the significance of the bit of the digital signal, the more speaklets that are energized in response to that bit. Thus, a time-varying sound level is generated by time-varying the number of speaklets emitting clicks. Louder sound is generated by increasing the number of speaklets emitting clicks. The present invention represents a substantial advance over the prior art in that sound is generated directly from a digital signal without the need to convert the digital signal first to an analog signal for driving a diaphragm.

WO 03/017717 A2



Published:

— without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD AND APPARATUS FOR RECONSTRUCTION OF SOUNDWAVES FROM DIGITAL SIGNALS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present invention claims priority based on U.S. Provisional Patent Application, Ser. No. 60/313,379 filed 17 August 2001 entitled "DIRECT DIGITAL EARPHONES", which is hereby incorporated by reference.

FIELD OF THE INVENTION

[0002] The present invention relates generally to the generation of a sound waveform directly from a digital signal and, more particularly, to the digital reconstruction of a sound waveform by providing a digital signal directly to microelectromechanical system (MEMS) devices.

BACKGROUND

[0003] Typical audio speakers use a vibrating diaphragm to produce soundwaves. The diaphragm is usually connected to a voice coil (i.e., an electromagnet). The voice coil is placed within the magnetic field of a permanent magnet. When an analog electrical signal is applied to the voice coil, the voice coil is either attracted to or repulse by the permanent magnet, depending on the polarity of the analog electrical signal. The analog electrical signal's alternating polarity imparts motion to the attached diaphragm, thus creating a soundwave. By varying the strength and the time it takes the analog electrical signal to change polarity, the volume and frequency, respectively, of the soundwave produced is regulated.

[0004] Most of today's sound recordings (for example, music, movies, etc.) are digitally recorded on, for example, CD's, DVD's, etc. Typical audio speakers, however, require that the digital sound recording be converted into an analog signal to drive the audio speaker's voice coil. Thus, additional digital-to-analog circuitry must be provided in the driver device (e.g., CD player, DVD player, etc.). The additional circuitry increases the complexity, size, cost, and power consumption of the driver device.

[0005] Thus, a need exists for a method and apparatus for directly reconstructing sound with a digital signal (i.e., without the need for converting the digital signal to an analog signal).

SUMMARY

[0006] The present invention is directed to the generation of sound by the superposition of discrete digital sound pulses from arrays of micromachined membranes called speaklets. The digital sound reconstruction (DSR) of the present invention is unlike any other reconstruction approach that has been demonstrated in that it offers true, digital reconstruction of sound directly from the digital signal. Traditional sound reconstruction techniques use a single to a few analog speaker diaphragms with motions that are proportional to the sound being created. In DSR, each speaklet produces a stream of clicks (discrete pulses of acoustic energy) that are summed to generate the desired sound waveform. With DSR, louder sound is not generated by greater motion of a diaphragm, but rather by a greater number of speaklets emitting clicks. Summarily, the time-varying sound level is not generated by a time-varying diaphragm motion, but rather by time-varying numbers of speaklets emitting clicks.

[0007] The present invention represents a substantial advance over the prior art in that sound is generated directly from a digital signal without the need to convert the digital signal first to an analog signal for driving a diaphragm. The elimination of the digital to analog circuitry reduces cost and nonlinearities resulting from such electronics. Furthermore, in the preferred method, the speaklets are produced using CMOS process techniques, which are well known and widely available. As a result, the speaklets can be produced in a uniform, cost effective manner. Those advantages and benefits, and others, will be apparent from the Detailed Description appearing below.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] To enable the present invention to be easily understood and readily practiced, the present invention will now be described for purposes of illustration and not limitation, in connection with the following figures wherein:

[0009] FIGS. 1A, 1B and 1C illustrate an idealized sound pulse (click) generated by a single speaklet's binary motion, a top view of an array at three different points in time, and a soundwave generated by the array, respectively.

[0010] FIG. 2 is a photograph of a 3-bit (7 speaklet) DSR earphone assembled and bonded in a TO-8 package. Under the chips, ventilation holes have been drilled through the package. Unused holes are filled to prevent air leakage.

[0011] FIG. 3A illustrates a 200 μ s long, 90 volt input pulse and the resulting response from one speaklet of FIG. 2.

[0012] FIG. 3B are curves illustrating the responses of the six other speaklets of FIG. 2 from the same pulse.

[0013] FIG. 4 illustrates the acoustic response of two individual speaklets and the additive response of both speaklets. The measured response is within 3% of the predicted additive response (mathematical sum of both speaklet responses).

[0014] FIG's 5A – 5C illustrates oscilloscope traces comparing the digital, acoustic reconstruction of a 500 Hz signal using a 1-bit, 2-bit, and 3-bit quantization, respectively.

[0015] FIG. 6 illustrates an embodiment of the present invention in which the number of speaklets activated is responsive to the position of each bit in a digital signal.

[0016] FIG. 7 illustrates an embodiment of the present invention in which the position of each bit in a digital signal determines which speaklet (of different sized speaklets) is to be activated, and

[0017] FIG. 8 illustrates a hybrid implementation in which certain parts of the digital signal are reproduced with a traditional speaker while other parts of the digital signal are reproduced with the apparatus of the present invention.

DETAILED DESCRIPTION

[0018] FIG. 1A illustrates an idealized sound pulse (click) generated by a single speaklet's binary motion. FIG. 1B is a top view of an array at three different points in time, i.e., as time t1, time t4, and time t6. At time t1, four speaklets have been activated. At time t4, no speaklets have been activated while at time t6 one speaklet has been activated. FIG. 1C illustrates how the clicks of FIG. 1A produced by the array of FIG. 1B are additive. Thus, the soundwave illustrated in FIG. 1C has a magnitude at time t1 equal to that of four clicks while the soundwave produced at time t4 has a magnitude of zero corresponding to the production of no clicks. The soundwave produced in FIG. 1C is produced directly from a digital signal. For example, the digital signal at time t1 has a value of "1 0", at time t4 a value of "0 0" and at time t6 a value of "0 1". Those values of the digital signal are used to directly drive speaklets without the need to convert the signal first into an analog signal. The digital sound reconstruction (DSR) of the present invention is unlike any other

reconstruction approach in that the digital signal is used to directly drive speaklets, producing clicks, which are summed to produce the output waveform. In DSR, each speaklet produces a stream of clicks to generate the desired soundwave. Thus, louder sound is not generated by greater motion of diaphragms, but rather by a greater number of speaklets emitting clicks. Similarly, the time varying sound level is not generated by a time-varying diaphragm motion, but rather by the time-varying numbers of speaklets emitting clicks.

[0019] In the current embodiment, the individual speaklets 16 are fabricated using CMOS-based processes as disclosed, for example, in International Publication No. WO 01/20948 A2 published 22 March 2001 and entitled "MEMS Digital-to-Acoustic Transducer with Error Cancellation", which is hereby incorporated by reference, although other methods of producing membranes may be used. For example, a serpentine metal and oxide mesh pattern (1.6 μm -wide beams and gaps) is repeated to form meshes with dimensions up to several millimeters. The mesh patterns are formed in a CMOS chip, etched, and released to form a suspended mesh, typically 10-50 μm above the substrate. A TeflonTM-like conformal polymer (0.5-1 μm) is then deposited onto the chip, covering the mesh and forming a membrane having an airtight seal over a cavity. Depending on the mesh geometry and gap between the membrane and substrate, a 50 – 90 volt potential is applied to electrostatically actuate the membrane. Ventilation holes are etched from the back, allowing greater movement of the membrane by decreasing the acoustic impedance on the membrane's backside and providing a mechanism for damping resonant oscillations. Each membrane forms a speaklet.

[0020] Test data for the present invention was obtained using an array 6 of seven speaklets 8 as shown in FIG. 2. The speaklets 8 measured 1.4 mm x 1.4 mm and were bonded to a TO-8 electronic package to construct a 3-bit digital earphone. Four of the seven speaklets 8 were electrically tied to the same input to form the most significant bit of sound, two speaklets 8 were tied to form the next most significant sound bit, and the remaining speaklet 8 formed the least significant bit. The earphone was connected to a Brüel and Kjaer (B&K) 4157 ear simulator and the earphone-microphone pair was put inside a B&K 4232 anechoic test chamber. FIG. 3A is a curve illustrating a 200 μsec long, 90 Volt input pulse and the acoustic output response of one speaklet 8 of FIG. 2. The responses of the other six speaklets 8 are shown in FIG. 3B and are

similar to that illustrated in FIG. 3A, the shape and amplitude of each differed slightly due to process variations across separate chips.

[0021] To demonstrate the additive nature of the acoustic responses, we measured the individual responses from a 200 μ sec 90 volt pulse for two speaklets. Then we drove both speaklets simultaneously with the same pulse and measured the collective response. As seen in FIG. 4, the measured collective response matches the predicted response within 3% at any point along the waveform.

[0022] FIG's 5A – 5C illustrate oscilloscope traces that measure the response of the device of FIG. 2 using a 1-bit, 2-bit and 3-bit quantization, respectively, of a 500 Hz sinusoid. The digital samples were regenerated at 20,000 samples/second using Labview 5.1 and a NI-6713 Data Acquisition (DAQ) card.

[0023] FIG. 6 is a simplified view of a digitally driven system 10 according to an embodiment of the present invention. The digitally driven system 10 is comprised of drive electronics 12 and an array 14 of speaklets 16. In the current embodiment, the speaklets 16 are microelectro mechanical system (MEMS) membranes. The array 14 is electrically connected to the drive electronics 12 via one or more leads 18.

[0024] Drive electronics 12 are operable to directly drive the speaklets 16 with a digital signal. The drive electronics 12 may, for example, be contained within a CD player, DVD player, MP3 player, etc. In the current embodiment, the digital signal is a multi-bit signal. For simplification (and not as a limitation), a 4-bit digital signal is used to illustrate the present invention in the current embodiment. It should be noted that digital signals having a different number of bits may be used (for example, 3-bit, 8-bit, 16-bit, 32-bit, etc.) while remaining within the scope of the present invention. It should be further noted that the term “directly drive” refers to activating a speaklet 16 without first converting the digital signal to an analog signal. Thus, in the current embodiment, digital-to-analog converters are not required.

[0025] In the embodiment of FIG. 6, array 14 is divided into four subsets (e.g., S1, S2, S3, S4). Each subset corresponds to one bit of the 4-bit signal. Each subset is comprised of one or more speaklets 16. More specifically, as illustrated in FIG. 6, subsets S1, S2, S3, and S4 are comprised of one, two, four, and eight speaklets 16, respectively. As illustrated, subset S4 represents the most significant bit of the 4-bit signal and subset S1 represents the least significant bit of the 4-bit signal. Drive electronics 12 are responsible for producing drive pulses for causing speaklets 16 to be

driven from their at rest position to their driven position whenever a “1” appears in the digital signal at the position associated with that set. The drive pulses therefore control the position of the membranes. For example, for the signal “0100” the speaklets of subset S3 are activated; for the signal “0110” the speaklets of subset S3, after returning to their at rest positions, are activated again along with the speaklets of subset S2. In that manner, a soundwave is directly reconstructed from the digital sound.

[0026] FIG. 7 illustrates an embodiment of the present invention in which the position of each bit in a digital signal determines which speaklet, from among a plurality of different sized speaklets, is to be activated. In FIG. 7, the speaklet corresponding to subset S2 is twice as large as the speaklet corresponding to subset S1. Similarly, the speaklet corresponding to subset S3 is twice as large as the speaklet for subset S2 and the speaklet corresponding to subset S4 is twice as large as the speaklet corresponding to subset S3. In such an embodiment, the signal “0100” would cause the speaklet S3 to be activated while the signal “0110” would cause the speaklet of subset S3, after returning to its rest position, to be activated again along with the speaklet corresponding to subset S2. Those of ordinary skill in the art will recognize that the embodiments of FIG. 6 and FIG. 7 may be combined. For example, the speaklet in FIG. 7 corresponding to subset S4 could be replaced with eight speaklets the size of the speaklet corresponding to subset S1 while leaving the size of the speaklets corresponding to subsets S1, S2 and S3 unchanged. Another example is for the speaklet of subset S3 to be comprised of two speaklets of the size of subset S2. In such an embodiment, the speaklet corresponding to subset S4 could be comprised of four speaklets of the size of the speaklet comprising subset S2. A wide variety of combinations can be obtained depending upon the process being used and limitations imposed by the layout. However, the effective sound producing area, resulting either from increased numbers of speaklets or speaklets of increased size for each set, except the set representative of the least significant bit, is twice that of the set representative of the preceding bit. For example, the set of speaklets for bit B1 is twice the number, or twice the size, of the set of speaklets for bit B0; the set of speaklets for bit B2 is twice the number, or twice the size, or some combination thereof, of the set of speaklets for bit B1, etc.

[0027] FIG. 8 illustrates yet another embodiment of the present invention. In FIG. 8, drive electronics 30 provide the four least significant bits, bits B0 – B3, to a digital-to-

analog converter 32 which is used to drive a conventional speaker 34. The remainder of the digital signal, the most significant bits B4 – B7, is used to drive arrays 36, 38, 40 and 42 which may be of the type illustrated in FIG. 6 or the type illustrated in FIG. 7, although the number of speaklets has been reduced for purposes of illustration. For example B4 could drive sixteen speaklets, B5 thirty-two speaklets, etc. The number of arrays, one, two, three or four, that are fired in response to the most significant bits is a function of the volume setting. For example, the higher the volume, the more arrays that are fired in response to the most significant bits.

[0028] The apparatus of the present invention can be manufactured using mass-produceable, micromachining technology to create the array of speaklets having characteristics that are extremely uniform from one speaklet to the next. Furthermore, the mechanical speaklets can be integrated with the necessary signal processing, addressing and drive electronics as such signal processing, addressing and drive electronics may be manufactured using the same CMOS techniques used to manufacture the speaklets. Use of MEMS fabrication technology allows for low-cost manufacturing; the utilization of a multitude of identical speaklets provides linearity as the speaklets are as close to being identical as possible within the tolerances of the lithographic processes used. Another advantage of the present invention is the extremely flat frequency response due to the fact that the resonant frequencies of the speaklets are far above the audio range. Because of the close physical location of the speaklets, their individual contributions are summed through the addition of the soundwaves they produce.

[0029] The division of labor amongst speaklets does not correspond to frequency range as in the case of a woofer, midrange, tweeter set-up. Rather, the number of speaklets that are activated is proportional to the desired sound pressure and not the frequency to be produced. Off-axis changes in frequency response due to interference effects are believed to be minimal in an earphone design utilizing the present invention because the acoustic pathlength differences are smaller than the shortest soundwave lengths of interest. Another advantage of an earphone constructed using the present invention is the extremely small sound pressures needed for normal use. Use of CMOS process technology allows the production of an earphone having small feature size thereby providing geometry control and registration of the device within an ear canal.

[0030] It should be recognized that the above-described embodiments of the invention are intended to be illustrative only. Numerous alternative embodiments may be devised by those skilled in the art without departing from the scope of the following claims. For example, in an alternative embodiment, an array having 256 speaklets (e.g. for an 8-bit DSR) may be used, and additional arrays provided for increased volume. The size of the speaklets' membranes may also be reduced to minimize ringing and lower the drive voltages necessary to actuate the speaklets. Additionally, arrays may be fabricated on a single chip to reduce process variations and improve response uniformity.

What is claimed is:

1. A method of reconstructing a waveform from a multibit digital signal, comprising:
using at least certain of said bits of said multi-bit digital signal to drive a plurality of MEMS membranes.
2. The method of claim 1 additionally comprising organizing said plurality of MEMS membranes into groups in which each group is representative of one bit of the digital signal.
3. The method of claim 2 wherein each group of membranes, except the group representative of the least significant bit, has twice the number of membranes as the group representative of the preceding bit of the digital signal.
4. The method of claim 2 wherein each group consists of at least one membrane and wherein each membrane, except the membrane representative of the least significant bit, has twice the area of the membrane representative of the preceding bit of the digital signal.
5. The method of claim 2 wherein each group, except the group representative of the least significant bit, has twice the effective area as the group representative of the preceding bit of the digital signal.
6. The method of claim 1 additionally comprising inputting at least certain of said bits to drive electronics to produce drive pulses, and wherein said drive pulses are used to control the position of a plurality of MEMS membranes.
7. The method of claim 1 wherein said using at least certain of said bits of said multi-bit digital signal to drive a plurality of MEMS membranes comprises generating a plurality of discrete acoustic energy pulses with said plurality of MEMS membranes.

8. A sound reproduction apparatus, comprising:
 - a plurality of MEMS membranes; and
 - a plurality of drivers for driving said plurality of membranes in response to a digital signal.
9. The apparatus of claim 8 wherein said membranes are organized into subsets, and wherein each of said plurality of drivers is responsive to a bit of a digital signal to drive one of said subsets, each of said subsets, except the subset representative of the least significant bit, has twice the effective area as the subset representative of the preceding bit.
10. The apparatus of claim 9 wherein the effective area is doubled by doubling the number of said MEMS membranes.
11. The apparatus of claim 9 wherein the effective area is doubled by doubling the size of said MEMS membrane.
12. The apparatus of claim 9 wherein the effective area is doubled by the combination of increasing the number and increasing the size of said MEMS membranes.

1/7

FIG. 1A



FIG. 1B

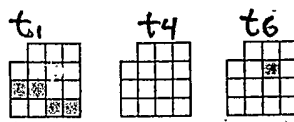


FIG. 1C

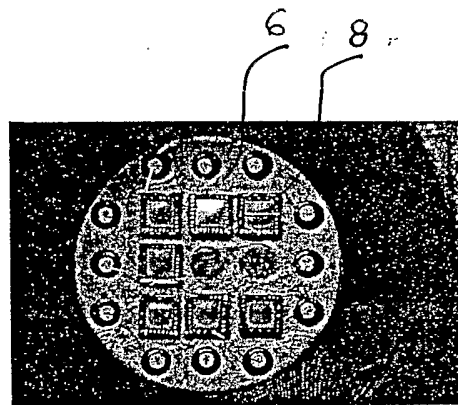
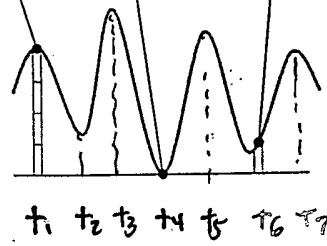


FIG. 2

2/7

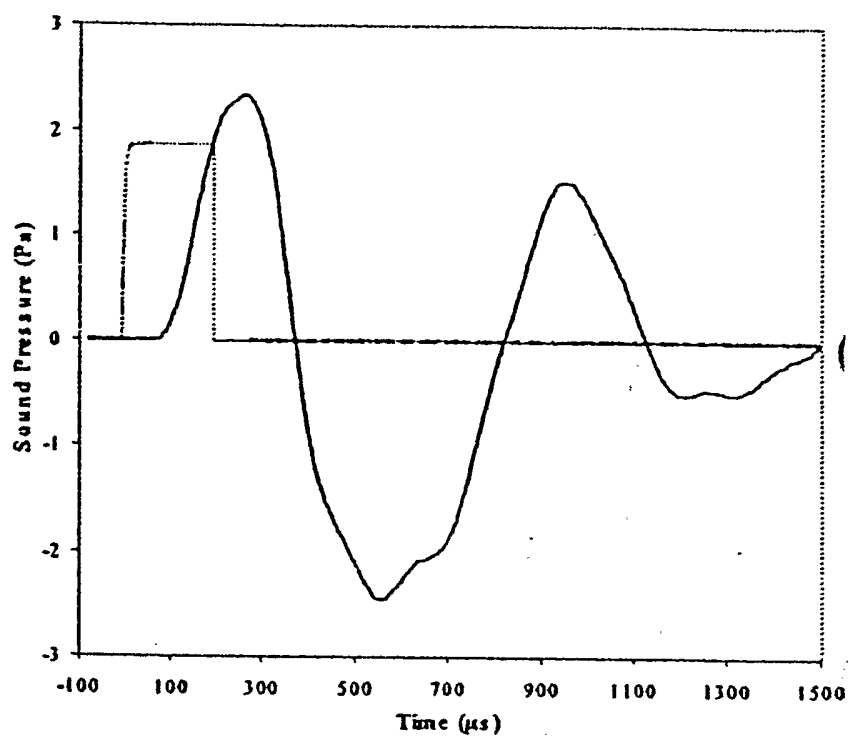


FIG. 3A

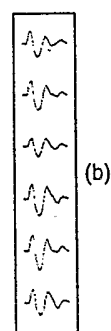


FIG. 3B

3/7

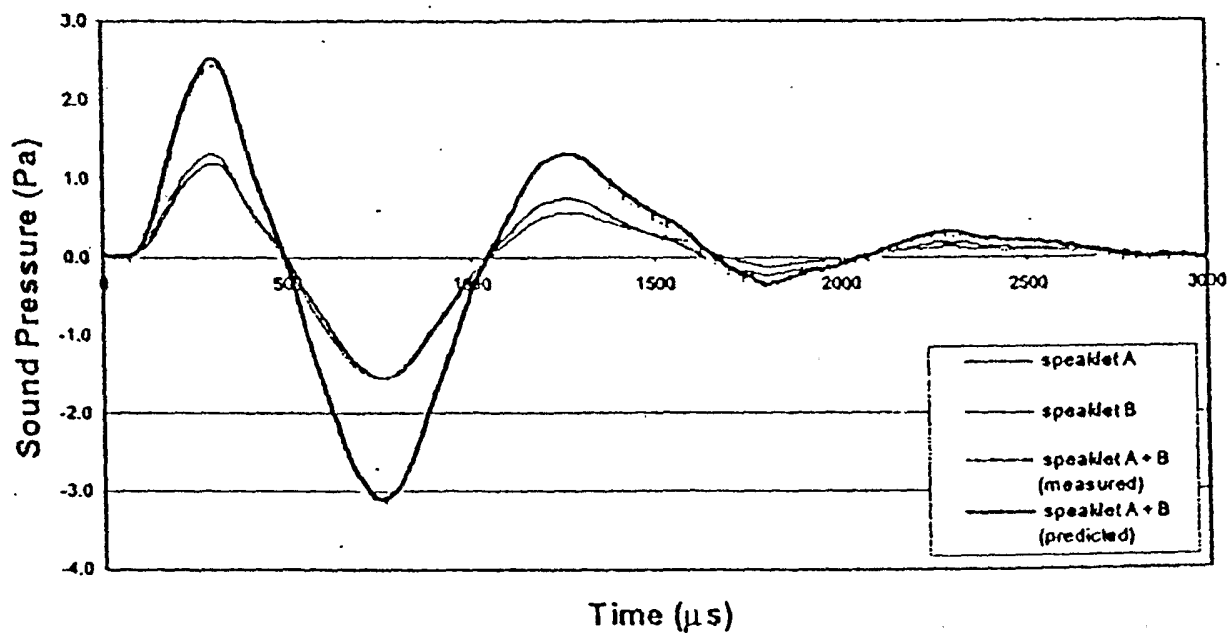
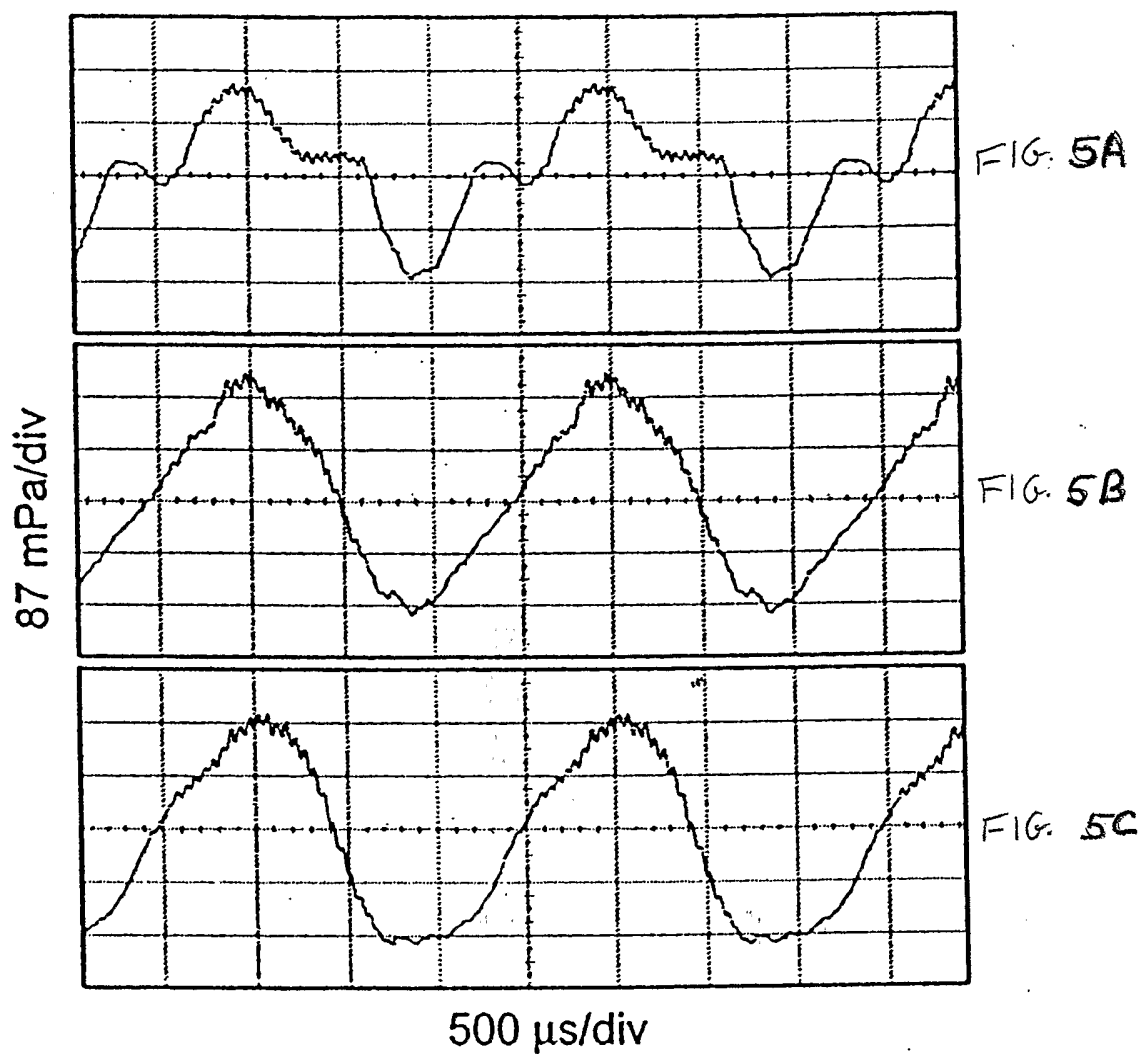


FIG. 4

4/7



5/7

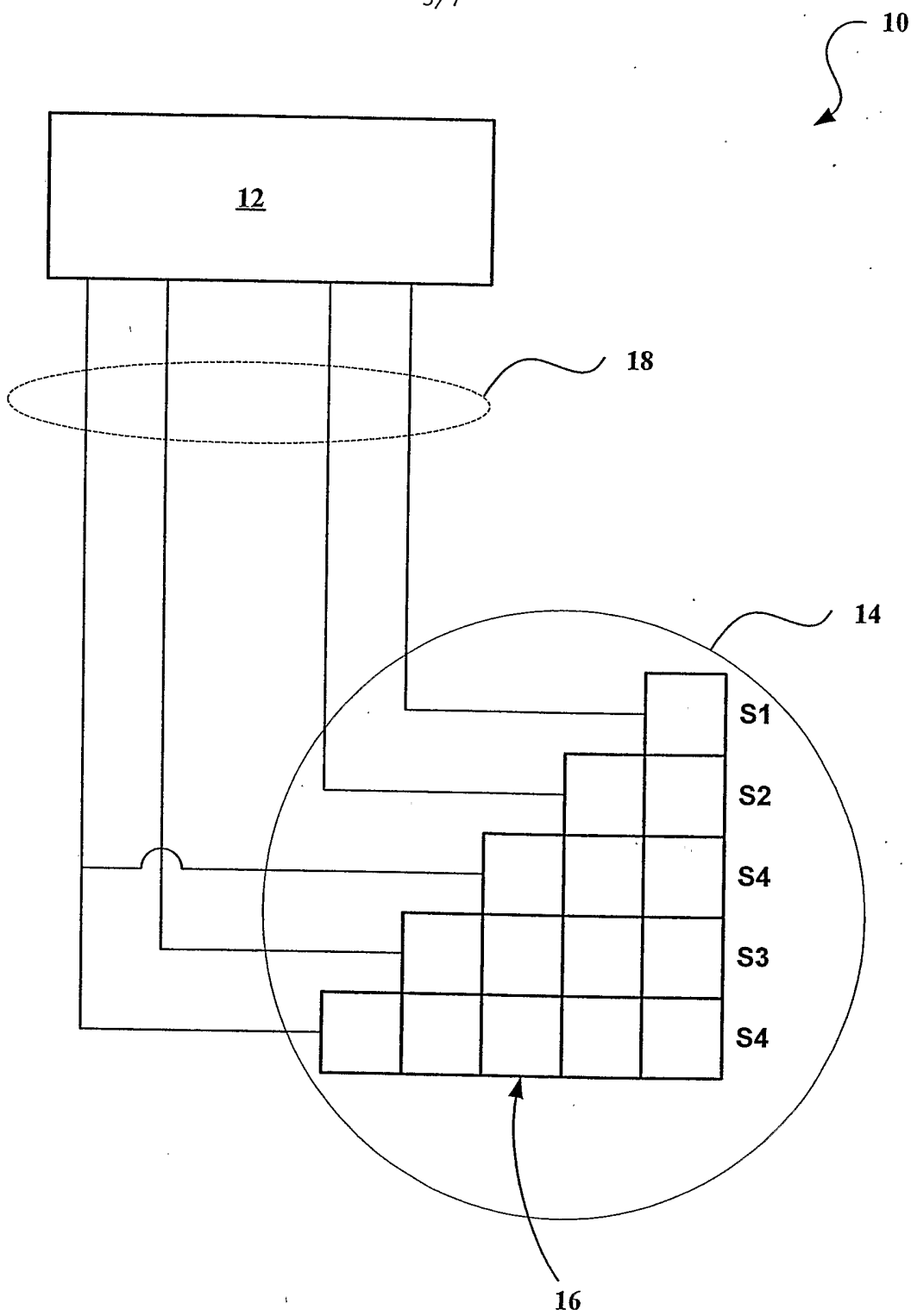


FIG 6

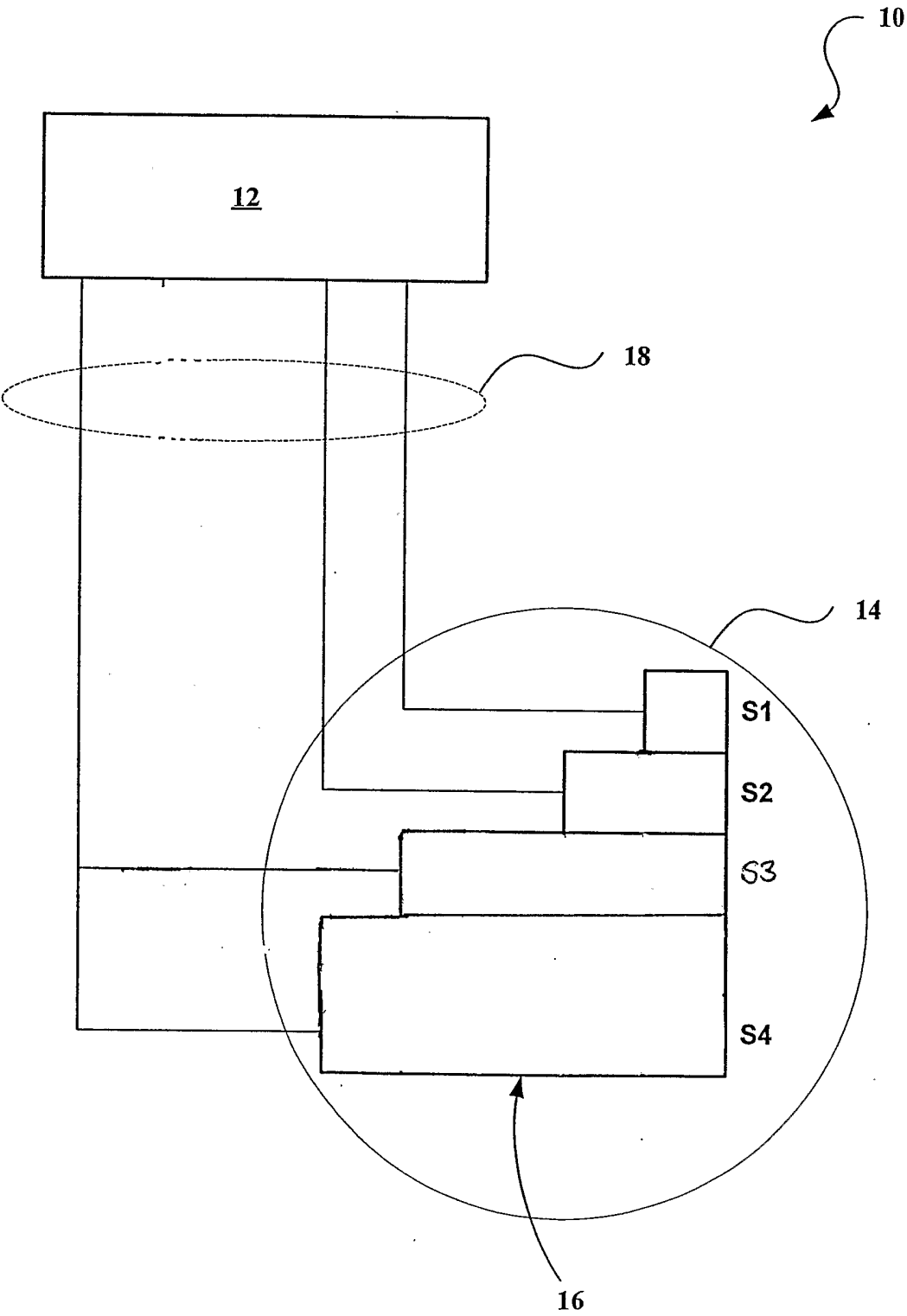


FIG. 7

7/7

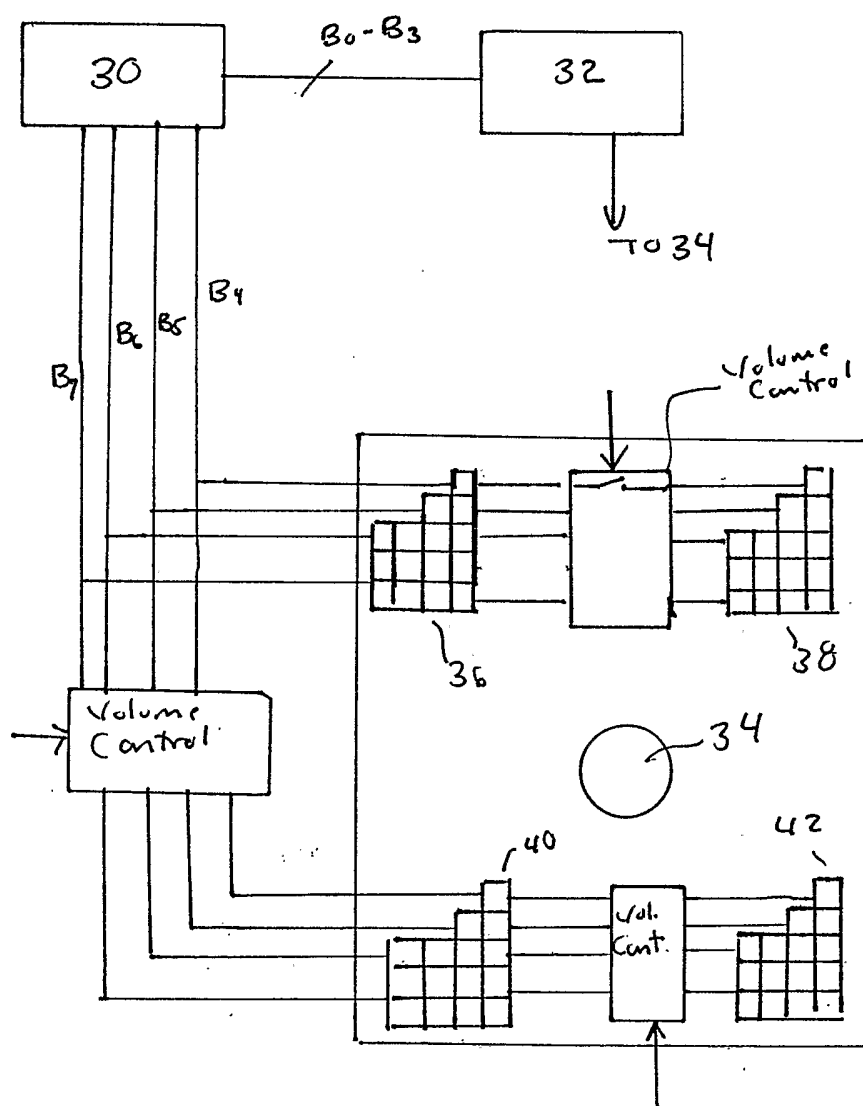


FIG. 8