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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 June 2016 (30.06.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/102954 A1

(51) International Patent Classification:
H04R 3/00 (2006.01) *G10L 19/04* (2013.01)
G10L 25/24 (2013.01)

(21) International Application Number:
PCT/GB2015/054122

(22) International Filing Date:
22 December 2015 (22.12.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/096,424 23 December 2014 (23.12.2014) US

(71) Applicant: **CIRRUS LOGIC INTERNATIONAL SEMICONDUCTOR LIMITED** [GB/GB]; Westfield House, 26 Westfield Road, Edinburgh, Central Scotland EH11 2QB (GB).

(72) Inventors: **LESSO, John Paul**; 8 Galachlaw Shot, Edinburgh, Lothian EH10 7JF (GB). **MELANSON, John Laurence**; 901 W. 9th St. #201, Austin, Texas 78703 (US).

(74) Agent: **O'CONNELL, David Christopher**; Haseltine Lake LLP, Redcliff Quay, 120 Redcliff Street, Bristol BS1 6HU (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: MICROPHONE UNIT COMPRISING INTEGRATED SPEECH ANALYSIS

(57) Abstract: A microphone unit has a transducer, for generating an electrical audio signal from a received acoustic signal; a speech coder, for obtaining compressed speech data from the audio signal; and a digital output, for supplying digital signals representing said compressed speech data. The speech coder may be a lossy speech coder, and may contain a bank of filters with centre frequencies that are non-uniformly spaced, for example mel frequencies.

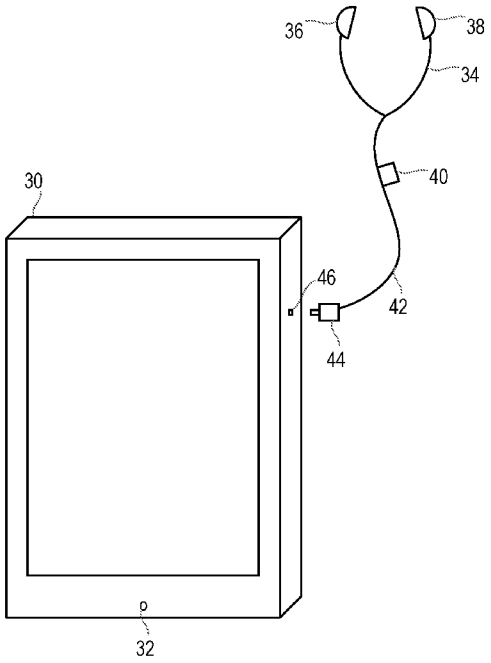


Figure 3

WO 2016/102954 A1

Published:

— *with international search report (Art. 21(3))*

MICROPHONE UNIT COMPRISING INTEGRATED SPEECH ANALYSIS

This disclosure relates to reducing the data bit rate on the interface of digital microphones, for example to minimise power consumption in Always-On Voice modes, while still passing enough information to allow downstream keyword detection or speech recognition functions.

BACKGROUND

Audio functionality is becoming increasingly prevalent in portable devices. Such functionality is present not only in devices such as phones that are reliant audio technology, but also in other wearable equipment or devices that may be controlled by voice, for instance voice-responsive toys such as listening-talking teddy bears. Such devices, including phones, will spend little of their time actually transmitting speech, yet one or possibly more microphones may be permanently enabled listening out for some voice command. Even a wearable accessory may be continuously on, awaiting a voice command, and will have little space for a battery, or may rely on some solar or mechanical energy harvesting, and so has severe power consumption requirements in a continuous standby mode as well as in a low-duty-cycle operating mode.

Microphone transducer and amplifier technology has improved, but generally a microphone package needs to drive its output signal some distance. Digital transmission offers advantages including noise immunity, but the usual formats for transmission of digital data from microphones are not particularly efficient in terms of signal line activity and the consequent power consumed in charging parasitic capacitances though a supply voltage at every logic level transition.

In a portable device such as a phone or tablet, containing one or more digital microphone, the digital microphone signal may have some distance to go from the microphone to a centralised smart codec chip or such, along a ribbon cable, or flex, or even across a densely populated printed circuit board. Even worse are applications where the microphone may be in a headset or earbuds or some acoustically desirable position on the user's clothing, distant from the handset or the main module of a distributed device.

However, even when largely otherwise inactive, there may be sophisticated signal processing to be performed, for example speaker recognition during voice-triggered wake-up, so solutions such as grossly degrading the resolution of the ADC therein may lead to unacceptable downstream processing results.

There is thus a requirement to reduce the power consumed in sending digital microphone data across a wired digital transmission link, while still conveying enough useful information in the transmitted signal to allow downstream function such as speech recognition.

Figure 1 illustrates a conventional digital microphone 10 communicating with a smart codec 22 in a host device 20, for example a phone, and Figure 2 illustrates the operating waveforms in a conventional digital microphone interface. A host device 20 transmits a clock CLK, typically at a frequency such as 3MHz, to the microphone 10, which uses this to clock an ADC 12 and to clock out from digital buffer interface Dout 14 a 1-bit oversampled delta-sigma stream DAT representing the acoustic signal input Px to the microphone transducer 16 providing the ADC input. Power is consumed in the system by the host 20 transmitting this clock signal CLK, and in particular by the microphone in sending a data stream DAT with an average 1.5MHz transition rate.

Power may be reduced by operating at a lower clock rate, say 768kHz, but this greatly increase the in-band quantisation noise or conversely restricts the usable bandwidth for a particular noise level. Even this only reduces the power by a factor of 4, so the power consumption is still significant, particularly in larger form factor devices or long cable runs.

Transmitting a delta-sigma stream is notably less efficient in terms of data bit rate and transition rate than transmitting a serial multi-bit pulse-code-modulated stream, but the latter generally requires an additional clock wire to transmit clocks to mark the start of each multi-bit word.

Secondly we note that an unfortunate side effect of reducing the delta-sigma sample clock rate may be to limit the bandwidth usable in terms of background quantisation noise to say 8kHz rather than say 20kHz. This may increase the word error rate (WER) for Voice Key Word detection (VKD). This may in turn lead to a higher incidence of

false positives and the system may spend more time in its awake mode thus significantly affecting the average complete system power consumption.

Additionally there is also a prevalent requirement for functions requiring even more accurate input audio data streams, such as speaker identification, as part of a voice-triggered wake-up function. It is known that using a wider bandwidth for the speaker identification captures more speech signal components and thus relaxes the need for high signal-to-noise (SNR) (e.g. relaxes the need for low acoustic background noise, or carefully optimised microphone placement) to get high enough accuracy for biometric purposes. Even in a high SNR environment a relatively wide signal bandwidth may improve speaker verification accuracy. This is at odds with the concept of reducing the frequency of the digital microphone clock to reduce the power consumption.

SUMMARY

According to a first aspect of the invention, there is provided a microphone unit, comprising:

a transducer, for generating an electrical audio signal from a received acoustic signal;

a speech coder, for obtaining compressed speech data from the audio signal; and

a digital output, for supplying digital signals representing said compressed speech data.

In embodiments of the invention, the microphone unit comprises a packaged microphone, for example a MEMS microphone, with on-chip or co-packaged integrated speech coder circuitry. This circuitry transmits data out of this package on a PCB trace or possibly headset cable to downstream circuitry that may perform more complex functions such as speech recognition, the transmitted data representing speech information coded in a speech-compressed format at a low bit rate to reduce the power consumed in physically transmitting the data.

In this disclosure, uncompressed data can be regarded as a numeric representation of samples in an uniformly sampled system, where the in-band signal is an approximation, in the audio band, of the audio input waveform, whereas compressed

data is typically derived from uncompressed data in such a way that the digital stream no longer directly represents the above, and has a lower bit rate.

Speech coding is an application of data compression of digital audio signals containing speech. Speech coding uses speech-specific parameter estimation using audio signal processing techniques to model the speech signal, and may be combined with generic data compression algorithms to represent the resulting modelled parameters in a compact bitstream.

Thus, compressed speech data may be [usually digital] data representing an audio signal in terms of speech-specific parameters calculated from the signal. For example, this may be the signal energy in a set of non-uniformly spaced frequency bins, or may use sub-band coding via say ADPCM of each sub-band. Data compression techniques may then be applied to these time-varying parameters, for example recoding scalars or vectors according to some codebook.

As examples, embodiments of the invention may use any speech compression standard, for example one using MDCT, MDCT-Hybrid subband, CELP, ACELP, Two-Stage Noise Feedback Coding (TSNFC), VSELP, RPE-LTP, LPC, Transform coding, or MLT, with suitable examples being AAC, AC-3, ALAC, ALS, AMBE, AMr, AMR-WB, AMR-WB+, apt-X, ATRAC, BroadVoice, CELT, Codec2, Enhanced AC-3, FLAC, any of the group of G.7xx standards, GSM-FR, iLBC, iSAC, Monkey's Audio, MP2, MP3, Musepack, Nellymoser Asao, Opus, Shorten, SILK, Siren 7, Speex, SVOPC, TTA, TwinVQ, Vorbis, WavPack, or Windows Media Audio.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 illustrates an audio processing system.

Figure 2 illustrates signals in the audio processing system of Figure 1.

Figure 3 illustrates a system, comprising a host device and an accessory.

Figure 4 illustrates an audio processing system.

Figure 5 illustrates a part of a microphone unit.

Figure 6 illustrates a part of a microphone unit.

Figure 7 illustrates a part of a microphone unit.

Figure 8 illustrates a compressive speech coder.

Figure 9 illustrates an audio processing system.

Figure 10 illustrates an audio processing system.

Figure 11 illustrates a part of a microphone unit in the audio processing system of Figure 10.

DETAILED DESCRIPTION

Figure 3 shows an audio system, as just one example of a system using the methods described herein.

Specifically, Figure 3 shows a device 30, which in this example takes the form of a smartphone or tablet computer. The methods described herein may be used with any device, but are described herein with reference to a particular example in which the device is a portable communications device. Thus, in this example, the host device 30 has audio processing capability.

Figure 3 shows an audio input 32, near which there is located a microphone, within the body of the device 30 and therefore not visible in Figure 3. In other devices, there may be multiple microphones. Figure 3 also shows an accessory device 34, which in this example takes the form of a pair of earphones, but which may be any device, in particular any audio accessory device. In this example, the pair of earphones has two earpieces 36, 38, each of which includes a speaker for reproducing sound in response to audio signals transferred from the host device 30. Each of the earpieces 36, 38 also includes at least one microphone, for example for detecting ambient noise in the vicinity of the wearer.

Signals representing the ambient sound are then transferred from the earphones to the host device 30. The host device may then perform various functions. For example, the host device may perform a noise cancellation function using an algorithm and generate anti-noise signals that it transfers to the earphones for playback. The effect of playing back the anti-noise signals is that the level of ambient noise heard by the wearer is reduced, and the wanted sounds (music, speech, or the like) that are also being transferred from the host device 30 are therefore more audible.

The accessory device 34 in this example also includes a microphone 40 that is positioned near to the user's mouth when wearing the earphones. The microphone 40 is suitable for detecting the user's speech. The accessory device 34 may be connected to the host device 30 by means of a cable 42. The cable 42 is detachable from at least one of the portable communications device and the audio accessory. In some embodiments, the cable 42 is permanently attached to the accessory device 34, and may be provided with a jack 44, to allow mechanical and electrical connection to or disconnection from the host device via a socket 46 provided on the host device. The cable may be in any suitable format.

The host device 30 includes circuitry for receiving signals from the microphone or microphones within the body of the device 30 and/or from the microphones in the earpieces 36, 38, and/or the microphone 40. The circuitry may for example include a codec 52, audio DSP, or other processing circuitry, which in turn may be connected to circuitry within the host device 30 such as an applications processor, and/or may be connected to a remote processor.

For example, the processing circuitry may be able to perform a speech processing function, such as recognising the presence of a trigger phrase in a speech input received by one or more of the microphones, identifying a speaker of the speech input, and/or recognising the content of a spoken command in order to be able to control the host device or another connected device on the basis of the user's spoken command.

Figure 4 shows an embodiment with a microphone unit 50 having a digital transmission format and method, for communication to a downstream smart codec 52, audio DSP, or other processing circuitry.

The microphone unit 50 comprises a transducer 54, an analogue-to-information converter (AIC) 56 and a digital output driver 58.

The analogue-to-information converter 56, or speech coder, or feature extraction block, may take several forms. It is well known that brute force digitisation of an audio signal is grossly inefficient in terms of the useful information conveyed, or usually required, as interpreted by say the human ear and brain or some machine equivalent. The basic concept is to extract features of the audio signal that may be particularly useful when interpreted downstream, as illustrated by the data stream Fx in Figure 4. A digital interface 58 then transmits a data stream FDAT carrying this coded speech signal to the codec 52. In one embodiment, a clock recognition block 60 in the codec 52 recovers some clock from the incoming data, and then a feature processing block 62 operates on the received feature information to perform functions such as voice activity detect or speaker recognition, delivering appropriate flags VDet to downstream processing circuitry or to control or configure some further or subsequent processing of its own. The codec 52 may comprise a clock generation circuit 66, or may receive a system clock from elsewhere in the host device.

Preferably the AIC 56 is asynchronous or self-timed in operation, so does not require a clock, and the data transmission may then also be asynchronous, as may at least early stages of processing of the feature data received by the codec. It may comprise an asynchronous ADC, for instance an Asynchronous Delta-Sigma Modulator (ADSM) followed by other analogue asynchronous circuitry or self-timed logic circuitry for digital signal processing.

However the microphone may generate its own clock if required by the chosen AIC circuit structure or FDAT data format.

In some embodiments the microphone unit may receive at least a low-frequency clock from the codec or elsewhere such as the system real time clock for use to synchronise or tune its internal clock generator using say locked-loop techniques. However, as will be discussed below, the feature data to be transmitted may typically be a frame produced at nominally say 30Hz or 10Hz, and the design of any speech processing function, say speech recognition, may have to accommodate a wide range of pitches and spoken word rate. Thus in contrast to the use case where music has to be recorded with accurate pitch and where any jitter may lead to unmusical

intermodulation, the clock in the voice recognition mode does not need an accurate or low-jitter sampling clock, so an on-chip uncalibrated low-power clock 64 may be more than adequate.

In some embodiments, the data may be transmitted as a frame or vector of data at some relatively high bit rate, such that there is a transitionless interval before the each next frame.

In all of the embodiments described herein, in which a microphone unit comprises a transducer and a feature extraction block, the transducer may comprise a MEMS microphone, with the MEMS microphone and the feature extraction block being provided in a single integrated circuit.

The microphone unit may comprise a packaged microphone, for example a MEMS microphone, with on-chip or co-packaged integrated speech coder circuitry or feature extraction block.

This speech coder circuitry, or feature extraction block, may transmit data out of the package on a PCB trace or possibly a cable such as a headset cable to downstream circuitry that may perform more complex functions such as speech recognition, the transmitted data representing speech information coded in a speech-compressed format at a low bit rate to reduce the power consumed in physically transmitting the data.

Figure 5 illustrates one embodiment of an AIC 56, in which the analog input signal is presented to an ADC 70, for example a 1-bit delta-sigma ADC, clocked by a sample clock CKM of nominally 768kHz. The delta-sigma data stream Dx is then passed to a Decimator, Window block, and a Framer 72 for decimating the data to a sample rate of say 16ks/s, suitable windowing and then framing for presentation to an FFT block 74 for deriving a set of Fourier coefficients representing the power (or magnitude) of the signal in each one of a set of equally spaced frequency bins. This spectral information is then passed through a mel-frequency filter bank 76 to provide estimates of the signal energy in each of a set of non-equally-spaced frequency bands.

This set of energy estimates itself may be used for output. Alternatively, each of these energy estimates is passed though a log block 78 to compand the estimate, and then

though a Discrete Cosine Transform block 80 to provide cepstral coefficients, known as Mel-frequency Cepstral Components (MFCC).

In one example the output cepstral coefficients comprise 15 channels of 12-bit words at a frame period of 30ms, thus reducing the data rate from the original 1-bit delta-sigma rate of 3Mbs/s or 786kb/s to 6kb/s.

Figure 6 illustrates another embodiment of an AIC 56, with some extra functional blocks in the signal path compared with Figure 5. In some other embodiments not all of these blocks may be present.

The analog input signal from the transducer element 90 is presented to an ADC 92, for example a 1-bit delta-sigma ADC, clocked by a sample clock CKM of nominally 768kHz generated by a local clock generator 94, which may be synchronised to a system 32kHz Real-Time Clock for instance, or which may be independent.

The delta-sigma data stream Dx is then decimated in a decimator 96 to a sample rate of, say 16ks/s.

It may then be passed to a pre-emphasis block 98 comprising a high-pass filter, to spectrally equalise the speech signal, most likely dominated by low-frequency components. This step may also be advantageous in reducing the effect of low-frequency background noise, for instance wind noise or mechanical acoustic background noise. There may also be a frequency-dependent noise reduction block at this point, as discussed below, to reduce noise in the frequency bands where it is most apparent.

The signal may then be passed to a windowing block 100, which may apply say a Hamming window or possibly some other windowing function to extract short-duration frames, say of time duration 10ms to 50ms, over each of which the speech may be considered stationary. The windowing block extracts a stream of short-duration frames by sliding the Hamming window along the speech signal by say half the frame length, or say sliding a 25ms window by 10ms, thus providing a frame of windowed data at a frame rate of 100 frames per second.

An FFT 102 block then performs a Fast Fourier Transform (FFT) on the set of windowed samples of each frame, providing a set of Fourier coefficients representing the power (or magnitude) of the signal in each one of a set of equally spaced frequency bins.

Each of these frame-by-frame sets of signal spectral components is then processed by a Mel-filter-bank 104, which maps and combines these linear-spaced spectral components onto frequency bins distributed to correspond more closely to the non-linear frequency sensitivity of the human ear, with a greater density of bins at low frequencies than at high frequencies. For instance, there may be 23 such bins, each with a triangular band-pass response, with the lowest frequency channel centred at 125Hz and spanning 125kHz while the highest frequency channel in centred at 3657Hz and spans 656Hz. In some embodiments, other numbers of channels or other non-linear frequency scales such as the Bark scale may be employed.

A log block 106 then applies a log scaling to the energy reported from each mel-frequency bin. This helps reduce the sensitivity to very loud or very quiet sounds, in a similar way to the non-linear amplitude sensitivity of human hearing.

The logarithmically compressed bin energies are then passed as a set of samples to a Discrete Cosine Transform block DCT 108 which applies a Discrete Cosine Transform on each set of logarithmically compressed bin energies. This serves to separate the slowly varying spectral envelope (or vocal tract) information from the faster varying speech excitation. The former is more useful in speech recognition, so the higher coefficients may be discarded. However is some embodiments these may be preserved, or possibly combined by weighted addition to provide at least some measure of energy for higher frequencies to aid in distinguishing sibilants or providing more clues for speaker identification. In some embodiments the higher order (3) coefficients may be generated in parallel with the lower ones.

The DCT block 108 may also provide further output data. For instance one component output may be the sum of all the log energies from each channel, though this may also be derived by a parallel total energy estimator EST 110 fed from un-pre-emphasised data.

There may also be a dynamic coefficient generator which may generate further coefficients based on the first-order or second-order frame-to-frame differences of the coefficients.

An equaliser (EQ) block 112 may adaptively equalise the various components relative to a flat spectrum, for instance using an LMS algorithm.

Before transmission, the data rate may be further reduced by a Data Compressor (DC) block 114, possibly exploiting redundancy or correlation between the coefficients expected due to the nature of speech signals. For example split vector quantisation to compress the MFCC vectors. In one example feature vectors of dimension 14 say may be split into pairs of sub vectors, each quantised to 5 or 6 bits say with a respective codebook at a frame period of 10ms. This may reduce the data rate to 4.4kb/s or lower, say 1.5kb/s if a 30ms frame period is used.

Additionally or alternatively, the data compressor may employ other standard data compression techniques.

Thus the data rate necessary to carry useful information concerning the speech content of the acoustic input signal has been reduced below that necessary for simple multi-bit or oversampled time-domain representations of the actual waveform by employing compression techniques at least in part reliant on known general characteristics of speech waveforms and of the human perception of speech, for instance in the use of non-linearly-spaced filter banks and logarithmic compression, or the separation of vocal tract information from excitation information referred to above. The outgoing data stream may be considered compressed speech data in that the output data has been compressed from the input signal in a manner particularly suitable for speech and for communication of the parameters of a speech waveform that convey information rather than general purpose techniques of signal digitisation and of compressing arbitrary data streams.

Having generated the compressed speech data, this data now needs to be physically transmitted to the codec or other downstream circuitry. In the case of an accessory connected to a host device by means of a cable (such as the headset 34 containing multiple microphones being connected to an audio device 30, as shown in Figure 1), the output data may be transmitted simply using two wires, one carrying the data (for

example 180 bits every 30ms, in the example of Figure 5), and the second carrying a sync pulse or edge every 30ms. The extra power of this low clock rate clock line is negligible compared to the already low power consumption of the data line. Similarly a two-wire link may be used between a microphone in the body of a device such as a mobile phone and a codec or similar on a circuit board inside the phone.

Standard data formats such as Soundwire™ or Slimbus™ may be used, or standard three-wire interfaces such as I2S.

Alternatively, , a one-wire serial interface may be employed, transmitting the data in a recurring predefined sequence of frames, in which a unique sync pattern may be sent at the start of every frame of words and recovered by simple and low power data and clock recovery circuitry in the destination device. The clock is preferably a low-power clock inside the microphone, whose precise frequency and jitter is unimportant since the feature data is nowhere near as clock critical as full-resolution PCM.

Nibbles of data may be sent using a pulse-length modulated (PLM) one-wire or two-wire format such as disclosed in published US Patent Application (US2013/0197920(A1)). The data may be sent with a sequence of pulses with a fixed leading edge, with the length of each pulse denoting the binary number. The fixed leading edge makes clock recovery simple.

Some slots in the outgoing data stream structure (PLM or non-PLM) may be reserved for identification or control functions. In this application, with continuous streams of data, occasional data bit errors may not have a serious impact. However, in some applications it may be desirable to protect at least the control data with some error-detection and/or correction scheme, e.g. based on Cyclic Redundancy Check bits embedded in the stream.

The speech coding to reduce the data rate and thus the average transition rate on the physical bus may thus greatly reduce the power consumption of the system. This power saving may be offset somewhat by the power consumed by the speech coding itself, but this processing may have otherwise had to be performed somewhere in the system in order to provide the keyword detection or speaker recognition or more general speech recognition function in any case. Also with decreasing transistor size

the power required to perform a given digital computation task is falling rapidly with time.

It is known that Mel-frequency Cepstral Component (MFCC) values are not very robust in the presence of additive noise. This may lead to false positives from a downstream voice keyword detector, which may lead to this block frequently triggering futile power-up of following circuitry, with a significant effect on average system power consumption.

In some embodiments the generation method may be modified, for instances by raising the log-mel-amplitudes (generated by the block 78 in the embodiment shown in Figure 5 or the block 106 in the embodiment shown in Figure 6) to a suitable power (around 2 or 3) before taking the DCT (in the block 80 in the embodiment shown in Figure 5 or the block 108 in the embodiment shown in Figure 6), which reduces the influence of low-energy components.

In some embodiments the parameters of the feature extraction may be modified according to a detected or estimated signal-to-noise ratio or other signal- or noise-related parameter associated with the input signal. For instance the number and centre-frequency of the cepstral frequency bins over which the mel-frequency energy is extracted may be modified.

In some embodiments, the cepstral coding block may comprise or be preceded by a noise reduction block for instance directly after a decimation block 72 or 96 or after a pre-emphasis block 98 that may already have removed some low frequency noise, or operating on the windowed framed data produced by block 100. This noise reduction block may be enabled when necessary by a noise detection block. The noise detection block may be analog and monitor the input signal Ax, or it may be digital and operate on the ADC output Dx. The noise detection block may flag when the level or spectrum or other characteristic of the received signal implies a high noise level or when the ratio of peak or average signal to noise falls below a threshold.

The noise reduction circuitry may act to filter the signal to suppress frequency bins where the noise, as monitored in time periods when there appears to be no voice as monitored by a Voice Activity Detector, is likely to exceed the signal at times where

there is a signal. For instance a Wiener Filter set up may be used to suppress noise on a frame-by-frame basis. The Wiener filter coefficients may be updated on a frame-by-frame basis and coefficient smoothed via a Mel-frequency filter bank followed by an Inverse Discrete Cosine Transform before application to the actual signal.

In some embodiments the Wiener noise reduction may comprise two stages. Each stage may incorporate some dynamic noise enhancement feature where the level of noise reduction performed is dependent on an estimated signal-to-noise ratio or other signal- or noise-related parameter or feature of the signal.

Various signal coding techniques where the output data transmitted is derived from the signal energies associated with each of a filter bank with non-uniformly spaced centre frequencies as described above, particularly cepstral feature extraction using MFCC coding, are compatible with many known downstream voice recognition or speaker recognition algorithms. In some cases the MFCC data may actually be forwarded from the codec, for example in an ETSI-standard MFCC form, for further signal processing either within the host device or transmitted to remote servers for processing “in the cloud”. This latter may reduce the data bandwidth required for transmission, and may be used to preserve speech quality in poor transmission conditions.

However in some embodiments the microphone may be required to deliver a more traditional output signal digitising the instantaneous input audio signal in say a 16-bit format at say 16ks/s or 48ks/s.

There may also be other applications in which some other format of signal is required. Traditionally this processing and re-formatting of the signal might take place within a phone applications processor or a smart codec with DSP capability. However given the presence of DSP circuitry in the microphone unit, necessary to reduce digital transmission power in stand-by or “Always-On” modes, this DSP circuitry may be usable to perform other speech coding methods in other use cases. As semiconductor manufacturing processes evolve with ever-decreasing feature sizes, and as the cost of each of these processes decreases over time with maturity, it becomes more feasible to actually integrate this functionality in the microphone unit itself, leaving any more powerful processing power elsewhere in the system freer to perform higher-level tasks. Or indeed in some end applications, the requirement for other signal-processing DSP may be removed, allowing perhaps some simpler non-DSP controller processor to be used.

Figure 7 illustrates a microphone unit 130 which may operate in a plurality of modes, with various degrees and methods of signal coding or compression. Thus, Figure 7 shows several different functional blocks. In some other embodiments, only a subset of these blocks is present.

The analog input signal from a transducer element 132 is presented to an ADC 134, for example a 1-bit delta-sigma ADC, and the resulting delta-sigma data stream Dx is then passed to one or more functional blocks, as described below.

The ADC may be clocked by a sample clock CKM, which may be generated by a local clock generator 136, or which may be received on a clock input 138 according to the operating mode.

The microphone unit may operate in a first, low power , mode in which it uses an internally generated clock and provides compressed speech data and a second, higher power, mode in which it receives an external clock and provides uncompressed data.

The operating mode may be controlled by the downstream control processor via signals received on a control input terminal 140. These inputs may be separate or may be provided by making the digital output line bi-directional. In some embodiments the operating mode may be determined autonomously by circuitry in the microphone unit. A control block 142 receives the control inputs and determines which of the functional blocks are to be active.

Thus, Figure 7 shows that the data stream Dx may be passed to a PDM formatting block 144, which allows the digitised time-domain output of the microphone to be output directly as a PDM stream. The output of the PDM formatting block 144 is passed to a multiplexer 146, operating under the control of the control block 142, and the multiplexer output is passed to a driver 148 for generating the digital output DAT.

Figure 7 also shows the data stream Dx being passed to a feature extraction block 150, for example for obtaining values based on using non-linear-spaced frequency bins, for instance MFCC values.

Figure 7 also shows the data stream Dx being passed to a compressive sampling block 152, for example for deriving a sparse representation of the incoming signal.

Figure 7 also shows the data stream Dx being passed to a lossy compression block 154, for example for performing Adaptive differential pulse-code modulation (ADPCM) or a similar form of coding.

Figure 7 also shows the data stream Dx being passed to a decimator 156.

In some embodiments, the data stream Dx is also passed to a lossless coding block to provide a suitable output data stream.

Figure 7 shows the outputs of the compressive sampling block 152, lossy compression block 154 and decimator 156 being connected to respective data buffer memory blocks 158, 160, 162. These allow the higher quality data generated by these blocks to be stored. Then, if analysis of a lower power data stream suggests that there is a need, power can be expended in transmitting the higher quality data for some further processing or examination that requires such higher quality data.

For example, analysis of a lower power data stream might suggest that the audio signal contains a trigger phrase being spoken by a recognized user of the device in a particular time period. In that case, it is possible to read from one of the buffer memory blocks the higher quality data relating to the same period of time, and to perform further analysis on that data, for example to confirm whether the trigger phrase was in fact spoken, or whether the trigger phrase was spoken by the recognized user, or performing more detailed keyword detection before awakening a greater part of a downstream system. Thus, the higher quality data can be used for downstream operations requiring better data, for example downstream speech recognition.

Figure 7 also shows the outputs of the feature extraction block 150, compressive audio processing block 152, and lossy compression block 154 being output through respective pulse length modulation (PLM) encoding blocks 164, 166, 168 and through the multiplexer 146, operating under the control of the control block 142, and the driver 148. Figure 7 also shows the output of the decimator 156 being output through a pulse

code modulation (PCM) encoding block 170 and through the multiplexer 146, operating under the control of the control block 142, and the driver 148.

The physical form of the transmitted output may differ according to what operating mode is selected. For instance high-data-rate modes may be transmitted using low-voltage-differential signalling for noise immunity, and the data scrambled to reduce emissions. On the other hand, in low-data rate modes, the signal may be low bandwidth and not so susceptible to noise and transmission line reflections and suchlike, and is preferably unterminated to save the power consumption associated with driving the termination resistance. In the lower power modes the signal swing, i.e. digital driver supply voltage, may be reduced.

Other operating parameters of the circuit may also be altered according to signal mode. For instance in low data rate modes the speed requirements of the DSP operations may be modest, and the circuitry may thus be operated at a lower logic supply voltage or divided master clock frequency than when performing more complex operations in conjunction with higher rate coding.

Although the AIC or feature-extraction based schemes above may provide particularly efficient methods of encoding and transmitting the essential information in the audio signal, there may be a requirement for the microphone unit to be able to operate also so as to provide a more conventional data format, say for processing by local circuitry or onward transmission for processing in the cloud, where such processing might not understand the more sophisticated signal representation, or where, for example, the current use case is for recording music in high quality.

In this case, it is advantageous for the initial conversion in ADC to be high quality, requiring a high quality low-jitter clock, and preferably synchronous with codec DSP main clock to avoid issues with sample-rate conversion to be synchronous to the codec master clock and/or the reference sample rate of the standard output digital PCM format. Thus, the microphone unit is operable in a first mode in which feature extraction and/or data compression is performed, and in a second mode, in which (for example) a clock is supplied from the codec, and the unit operates in a manner similar to that shown in Figure 1.

The digital microphone unit may thus be capable of operating in at least two modes – ADC (analog-digital conversion) or AIC (analog-information conversion). In ADC mode the PCM data from the ADC is transmitted, in AIC mode data extracted from the ADC output is coded, particularly for speech.

In further embodiments, the microphone unit is operable in one mode to perform lossy low-bit-rate PCM coding. For example, the unit may contain a lossy codec such as an ADPCM coder, with a sample rate that in some embodiments may be selectable, for example between 8ks/s – 24ks/s.

In some embodiments, the microphone unit has a coding block for performing μ-law and/or A-law coding, or coding to some other telephony standard. For example, the microphone unit in some embodiments has coding blocks for MDCT, MDCT-Hybrid subband, CELP, ACELP, Two-Stage Noise Feedback Coding (TSNFC), VSELP, RPE-LTP, LPC, Transform coding, or MLT coding.

In other embodiments, the microphone unit is operable in a mode in which it outputs compressive sampled PCM data, or any scheme that exploits signal sparseness.

Figure 8 illustrates an embodiment of a compressive speech coder that may be used in any of the embodiments described or illustrated herein. The output of an ADC 190 is passed through a decimator 192 to provide (for example) 12 bit data at 16ks/s or 48ks/s. This data is sampled at an average sample rate of only say 48Hz or 1kHz but with a sampling time randomised by a suitable random number generator or random pulse generator 194.

Thus, the sampling circuit samples the input signal at a sample rate less than the input signal bandwidth, and the sampling instants are caused to be distributed randomly in time.

Figure 9 shows a system using such a compressive speech coder. Thus, the microphone unit 200, including a compressive ADC 202, is connected to supply very low data rate data to a codec 204. With the aid of prior knowledge of the signal

statistics, downstream circuitry 206 may either perform a partial reconstruction (computationally cheap) to do sparse feature extraction in a low power mode, or a full reconstruction (computationally more expensive) to get Nyquist type voice for onwards transmission. Note there are known post-processing algorithm blocks, such as the block 208, for performing "sparse recognition" that are compatible with such compressive sampling formats. In such algorithms the sparse representation of a signal is matched to a linear combination of a few atoms from a pre-defined dictionary, which atoms may be obtained a priori by using machine learning techniques to learn an over-complete dictionary of primary signals (atoms) directly from data, so that the most relevant properties of the signals may be efficiently captured.

Sparse extraction may have some benefits in performance of feature extraction in the presence of noise. The noise is not recognised as comprising any atom component, so does not appear in the encoded data. Such ignorance of input noise may thus avoid unnecessary activation of downstream circuitry and avoid the power consumption increasing in noisy environments relative to quiet environments.

Figure 10 shows an embodiment in which a microphone unit 210 is connected to supply very low data rate data to a codec 212, and in which, to further reduce the power consumption, some, if not all, of the feature extraction is performed using Analogue Signal Processing (ASP). This a signal from a microphone transducer is passed to an analog signal processor 214, and then to one or more analog-to-digital converter 216, and then to an optional digital signal processor 218 in the microphone unit 210. Feature recognition 220 is then performed in the codec 212.

Figure 11 shows in more detail the processing inside an embodiment of the microphone unit 210, in which a large part of the signal processing is performed by analogue rather than digital circuitry. Thus, an input signal is passed through a plurality of band pass filters (three being shown in Figure 11 purely by way of illustration) 240, 242, 246. The band pass filters are constant Q and equally spaced in mel frequency. The outputs are passed to log function blocks 248, 250, 252, which may be achieved using standard analogue design techniques based for instance on applying the input signal via a voltage-to-current converted signal into an I-V two-port circuit with a logarithmic current-to-voltage conversion such as a semiconductor diode. The outputs are passed to a plurality of parallel ADCs 252, 254, 256. The ADCs may comprise voltage-controlled oscillators, whose frequency is used as a representation of their

respective input signals. These are simple and low power, and their linearity is not important in this application. These simple ADCs may have significantly reduced power and area even in total compared to the main ADC. The state of the art for similar circuit blocks in say the field of artificial cochleas is below 20 microwatts.

In all of the embodiments described herein, the microphone, ADC, and the speech coding circuitry may advantageously be located close together, to reduce high-data-rate signal paths of the digital data before data-rate reduction. All three components may be packaged together. At least two of these three components may be co-integrated on an integrated circuit.

The microphone may be a MEMS transducer, which may be capacitive, piezo-electric or piezo resistive, and co-integrated with at least the ADC.

The skilled person will recognise that various embodiments of the above-described apparatus and methods may be, at least partly, implemented using programmable components rather than dedicated hardwired components. Thus embodiments of the apparatus and methods may be, at least partly embodied as processor control code, for example on a non transitory carrier medium such as a disk, CD- or DVD-ROM, programmed memory such as read only memory (Firmware), or on a data carrier such as an optical or electrical signal carrier. In some applications, embodiments of the invention may be implemented, at least partly, by a DSP (Digital Signal Processor), ASIC (Application Specific Integrated Circuit) or FPGA (Field Programmable Gate Array). Thus the code may comprise conventional program code or microcode or, for example code for setting up or controlling an ASIC or FPGA. The code may also comprise code for dynamically configuring re-configurable apparatus such as re-programmable logic gate arrays. Similarly the code may comprise code for a hardware description language such as Verilog™ or VHDL (Very high speed integrated circuit Hardware Description Language). As the skilled person will appreciate, the code may be distributed between a plurality of coupled components in communication with one another. Where appropriate, the embodiments may also be implemented using code running on a field-(re-)programmable analogue array or similar device in order to configure analogue hardware.

It should be understood—especially by those having ordinary skill in the art with the benefit of this disclosure—that that the various operations described herein, particularly in connection with the figures, may be implemented by other circuitry or other hardware components. The order in which each operation of a given method is performed may be changed, and various elements of the systems illustrated herein may be added, reordered, combined, omitted, modified, etc. It is intended that this disclosure embrace all such modifications and changes and, accordingly, the above description should be regarded in an illustrative rather than a restrictive sense.

Similarly, although this disclosure makes reference to specific embodiments, certain modifications and changes can be made to those embodiments without departing from the scope and coverage of this disclosure. Moreover, any benefits, advantages, or solutions to problems that are described herein with regard to specific embodiments are not intended to be construed as a critical, required, or essential feature or element.

Further embodiments likewise, with the benefit of this disclosure, will be apparent to those having ordinary skill in the art, and such embodiments should be deemed as being encompassed herein.

CLAIMS

1. A microphone unit, comprising:

a transducer, for generating an electrical audio signal from a received acoustic signal;

a speech coder, for obtaining compressed speech data from the audio signal; and

a digital output, for supplying digital signals representing said compressed speech data.

2. A microphone unit as claimed in claim 1, wherein the speech coder contains a bank of filters with non-uniformly spaced centre frequencies.

3. A microphone unit as claimed in claim 2, wherein the centre frequencies are mel frequencies.

4. A microphone unit as claimed in claim 2 or 3, wherein outputs of the bank of filters are coupled to a log weighting block and a discrete cosine transform block to provide cepstral coefficients.

5. A microphone unit as claimed in one of claims 1 to 4, operable in a mode in which the digital output supplies uncompressed speech data.

6. A microphone unit as claimed in claim 1, comprising a compressive sampling coder comprising a sampling circuit which samples the input signal at a sample rate less than the input signal bandwidth, wherein the sampling instants are caused to be distributed randomly in time.

7. A microphone unit as claimed in one of claims 1 to 6, wherein the speech coder is a lossy speech coder.

8. A microphone unit as claimed in claim 7, wherein the lossy speech coder uses at least one coding technique selected from ADPCM, MDCT, MDCT-Hybrid subband, CELP, ACELP, Two-Stage Noise Feedback Coding (TSNFC), VSELP, RPE-LTP, LPC, Transform coding, and MLT.

9. A microphone unit as claimed in claimed in one of claims 1 to 8, wherein the speech coder and the digital output are provided on a single integrated circuit.

10. A microphone unit as claimed in claim 9, wherein the transducer is provided on said integrated circuit.

11. A microphone unit as claimed in claim 10, wherein the transducer comprises a MEMS microphone.

12. A microphone unit as claimed in one of claims 1 to 11, wherein the transducer is for generating an analog audio signal,

further comprising an analog-to-digital converter, for generating a digital audio signal from the analog audio signal, wherein the speech coder is connected to the analog-to-digital converter, for obtaining speech feature values from the digital audio output signal.

13. A microphone unit as claimed in one of claims 1 to 12, further comprising data compression circuitry, for receiving said cepstral feature values from the cepstral feature extraction block, and for generating reduced bit rate signals for supply to said digital output.

14. A microphone unit as claimed in claim 13,

wherein said reduced bit rate signals contain data at 20k bits/second or lower.

15. A microphone unit as claimed in one of claims 1 to 14, wherein the digital output is configurable in a first mode, for supplying said digital signals representing said compressed speech data, and in a second mode, for supplying digital signals representing time samples of said audio signal.

16. A microphone unit as claimed in claim 15, wherein the digital signals representing time samples of said audio signal are supplied at a higher data rate than the digital signals representing said compressed speech data.

17. A microphone unit as claimed in claim 16, wherein the digital signals representing time samples of said audio signal are PCM signals supplied at a sample rate of 16kHz or higher.

18. A microphone unit as claimed in claim 16, wherein the digital signals representing time samples of said audio signal are PDM signals at a sample rate of 0.5MHz or higher.

19. A microphone unit as claimed in one of claims 15 to 18, wherein the digital output is configurable between the first mode and the second mode based on a command received from a separate device to which the digital output is connected.

20. A microphone unit as claimed in one of claims 15 to 18, wherein the digital output is configurable between the first mode and the second mode based on a command generated in the microphone unit.

21. A microphone unit as claimed in claim 20, comprising a voice activity detector, for generating said command.

22. A microphone unit as claimed in claim 21, wherein said command causes the digital output to enter the first mode in response to detecting voice activity in the audio signal.

23. A microphone unit as claimed in one of claims 1 to 22, further comprising noise cancelling circuitry, for reducing the effects of ambient noise in the output digital signals.

24. A microphone unit as claimed in claim 23, wherein the noise cancelling circuitry comprises a Wiener filter.

25. A microphone unit as claimed in claim 23, wherein the noise cancelling circuitry comprises an adaptive Wiener filter.

26. A microphone unit, comprising:

a transducer, for generating an audio signal;

a bit rate compression block, for obtaining bit-rate compressed data from the audio signal; and

a digital output, for supplying digital signals representing said bit-rate compressed data.

27. A microphone unit as claimed in claim 26, wherein the transducer comprises a MEMS microphone.

28. A microphone unit, comprising:

a transducer, for generating an audio signal;

a speech feature extraction block, for obtaining speech feature values from the audio signal; and

a digital output, for supplying digital signals representing said speech feature values.

29. A microphone unit as claimed in claim 27, wherein the speech feature extraction block comprises means for obtaining speech feature values from the audio signal by means of a filter bank for providing estimates of the signal energy in each of a set of non-equally-spaced frequency bands.

30. A microphone unit as claimed in claim 29, wherein the non-equally-spaced frequency bands are mel frequencies.

31. A microphone unit as claimed in one of claims 28 to 30, wherein the speech feature extraction block comprises means for obtaining cepstral features.

32. A microphone unit as claimed in one of claims 28 to 31, wherein the speech feature extraction block and the digital output are provided on a single integrated circuit.

33. A microphone unit as claimed in claim 32, wherein the transducer is provided on said integrated circuit.

34. A microphone unit as claimed in claim 33, wherein the transducer comprises a MEMS microphone.

35. A microphone unit as claimed in one of claims 28 to 34,

wherein the transducer is for generating an analog audio signal,

further comprising an analog-to-digital converter, for generating a digital audio signal from the analog audio signal, wherein the speech feature extraction block is

connected to the analog-to-digital converter, for obtaining speech feature values from the digital audio output signal.

36. A microphone unit as claimed in claim 35 when dependent on one of claims 32 to 34, wherein the analog-to-digital converter is provided on said integrated circuit.

37. A microphone unit as claimed in one of claims 28 to 36, further comprising data compression circuitry, for receiving said speech feature values from the speech feature extraction block, and for generating reduced bit rate signals for supply to said digital output.

38. A microphone unit as claimed in claim 37, wherein the data compression circuitry operates using a predetermined code book.

39. A microphone unit as claimed in claim 37, wherein the data compression circuitry performs multi-vector coding.

40. A microphone unit as claimed in one of claims 28 to 39, wherein said digital signals contain data at 20k bits/second or lower.

41. A microphone unit as claimed in one of claims 28 to 40, wherein the digital output is configurable in a first mode, for supplying said digital signals representing said speech feature values, and in a second mode, for supplying digital signals representing time samples of said audio signal.

42. A microphone unit as claimed in one of claims 28 to 40, wherein the digital output is configurable in a first mode, for supplying said digital signals representing a data compressed version of said speech feature values, and in a second mode, for supplying digital signals representing time samples of said audio signal.

43. A microphone unit as claimed in claim 42, wherein said digital signals representing the data compressed version of said speech feature values comprise ADPCM compressed data.

44. A microphone unit as claimed in one of claims 41 to 43, wherein the digital signals representing time samples of said audio signal are supplied at a higher data rate than the digital signals representing said speech feature values.

45. A microphone unit as claimed in claim 44, wherein the digital signals representing time samples of said audio signal are PCM signals supplied at a sample rate of 16kHz or higher.

46. A microphone unit as claimed in claim 44 or 45, wherein the digital signals representing time samples of said audio signal are PDM signals at a sample rate of 0.5MHz or higher.

47. A microphone unit as claimed in one of claims 42 to 46, wherein the digital output is configurable between the first mode and the second mode based on a command received from a separate device to which the digital output is connected.

48. A microphone unit as claimed in one of claims 42 to 46, wherein the digital output is configurable between the first mode and the second mode based on a command generated in the microphone unit.

49. A microphone unit as claimed in claim 48, comprising a voice activity detector, for generating said command.

50. A microphone unit as claimed in claim 49, wherein said command causes the digital output to enter the first mode in response to detecting voice activity in the audio signal.

51. A microphone unit as claimed in one of claims 28 to 50, further comprising noise cancelling circuitry, for reducing the effects of ambient noise in the output digital signals.

52. A microphone unit as claimed in claim 51, wherein the noise cancelling circuitry comprises a Wiener filter.

53. A microphone unit as claimed in claim 52, wherein the noise cancelling circuitry comprises an adaptive Wiener filter.

54. A microphone unit as claimed in one of claims 28 to 53, wherein the speech feature extraction block comprises a plurality of analog filters providing signals in different respective frequency bands, and means for extracting features from said signals.

55. A microphone unit as claimed in claim 54, wherein the speech feature extraction block comprises a plurality of analog-to-digital converters connected to the outputs of the analog filters.

56. A microphone unit, comprising:

a transducer, for generating an audio output signal;

a feature extraction block, for deriving speech features from the audio output signal; and

a digital output, for driving digital data representing said speech features over a digital data line.

57. A microphone unit as claimed in claim 56, wherein the transducer comprises a MEMS microphone, and the MEMS microphone and the feature extraction block are provided in a single integrated circuit.

58. A microphone unit, comprising:

a transducer, for generating an audio output signal;

a feature extraction block, for deriving speech features from the audio output signal; and

a digital output, having a first mode for driving digital data representing said audio output over a digital data line at a first data rate, and having a second mode for driving digital data representing said derived features over a digital data line at a second data rate.

59. A microphone unit as claimed in claim 58, wherein the transducer comprises a MEMS microphone, and the MEMS microphone and the feature extraction block are provided in a single integrated circuit.

60. A microphone unit, comprising:

a transducer, for generating an audio output signal;

a digitiser, for generating a stream of digitised values representing said audio output signal;

a feature extraction block, for deriving speech features from the audio output signal; and

a digital output, having a first mode for driving said stream of digitised values over a digital data line at a first data rate, and having a second mode for driving digital data representing said derived features over a digital data line at a second data rate.

61. A microphone unit as claimed in claim 60, wherein the transducer comprises a MEMS microphone, and the MEMS microphone and the feature extraction block are provided in a single integrated circuit.

62. A digital microphone capable of operating in two modes:

a first ADC mode where the output is a digital version of the analogue input signal; and

a second AIC mode where the output is coded using the fact that input signal has a sparse representation in a given domain.

63. A portable device, having a microphone unit as claimed in any of claims 1 to 61 connectable thereto, the portable device comprising voice processing circuitry for receiving said digital signals representing said speech feature values.

64. A portable device as claimed in claim 63, wherein the voice processing circuitry is adapted for performing voice processing on the received digital signals.

65. A portable device as claimed in claim 64, wherein the voice processing circuitry comprises speech recognition circuitry for recognizing a speaker.

66. A portable device as claimed in claim 64 or 65, wherein the voice processing circuitry comprises speech recognition circuitry for recognizing spoken words in an input.

67. A portable device as claimed in claim 63, wherein the voice processing circuitry is adapted for transmitting data contained in said received digital signals to a remote processor.

68. A portable device as claimed in claim 67, wherein the voice processing circuitry is adapted for transmitting said data to the remote processor in an ETSI-standard MFCC form.

69. A portable device as claimed in one of claims 63 to 68, wherein the microphone unit as claimed in one of claims 1 to 61 is connectable to the portable device by a wired connection.

70. A portable device as claimed in one of claims 63 to 69, wherein the microphone unit as claimed in one of claims 1 to 61 is provided in a headset.

71. A portable device as claimed in one of claims 63 to 70, wherein the portable device comprises a portable communications device, a computing device, a wearable device, a games device, or a noise cancelling device such as a NC headset.

72. An audio processing system, comprising:

a microphone unit and an audio processing unit, wherein the microphone unit and the audio processing unit are connectable by means of a wired connection,

wherein the microphone unit comprises:

a transducer, for generating an audio output signal;

a feature extraction block, for deriving speech features from the audio output signal; and

a digital output, for driving digital data representing said speech features over the wired connection.

73. An audio processing system as claimed in claim 72, wherein the transducer comprises a MEMS microphone, and the MEMS microphone and the feature extraction block are provided in a single integrated circuit.

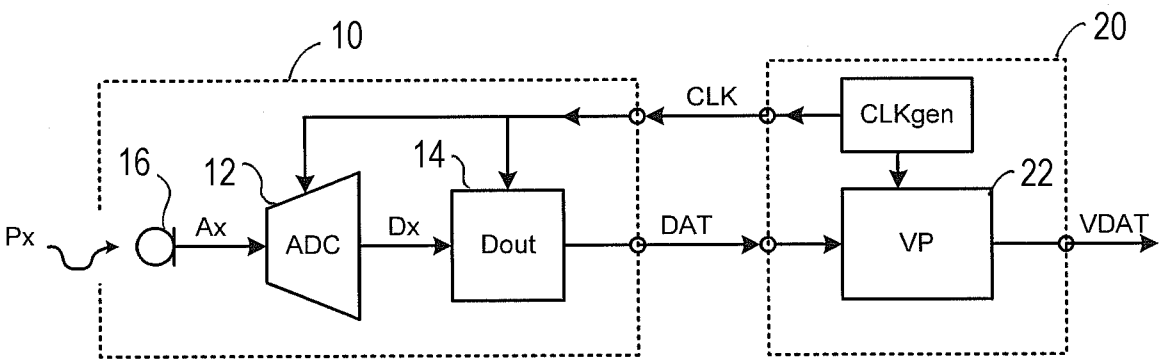


Figure 1

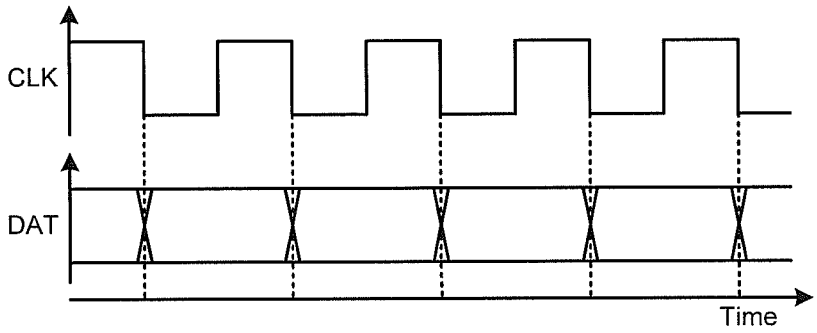


Figure 2

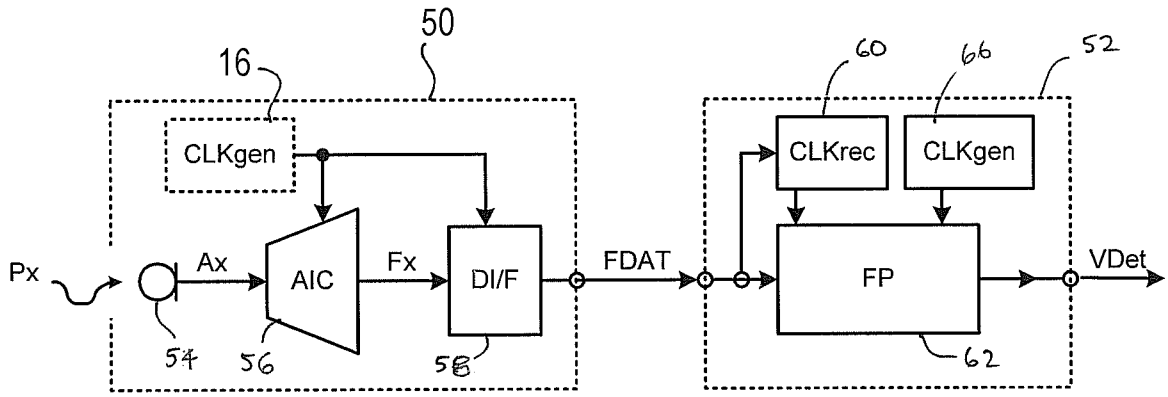


Figure 4

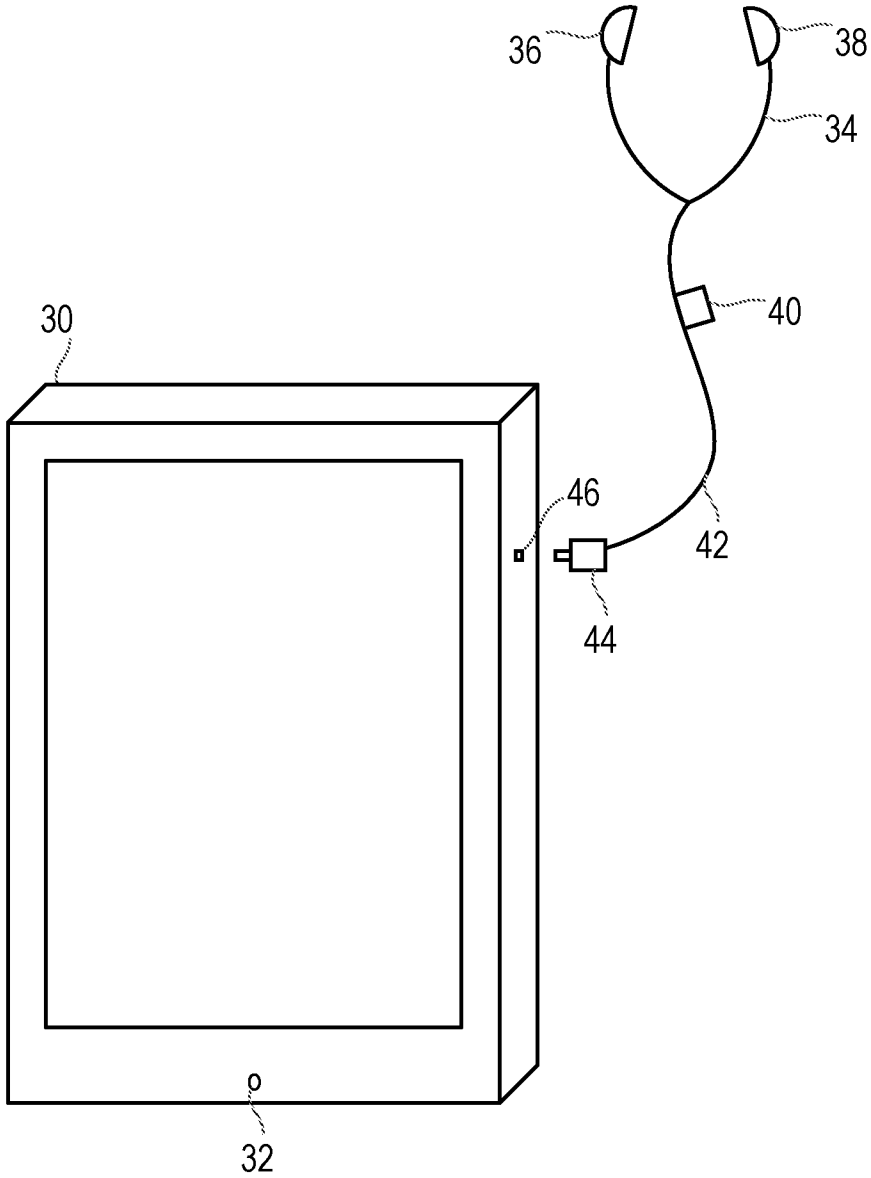


Figure 3

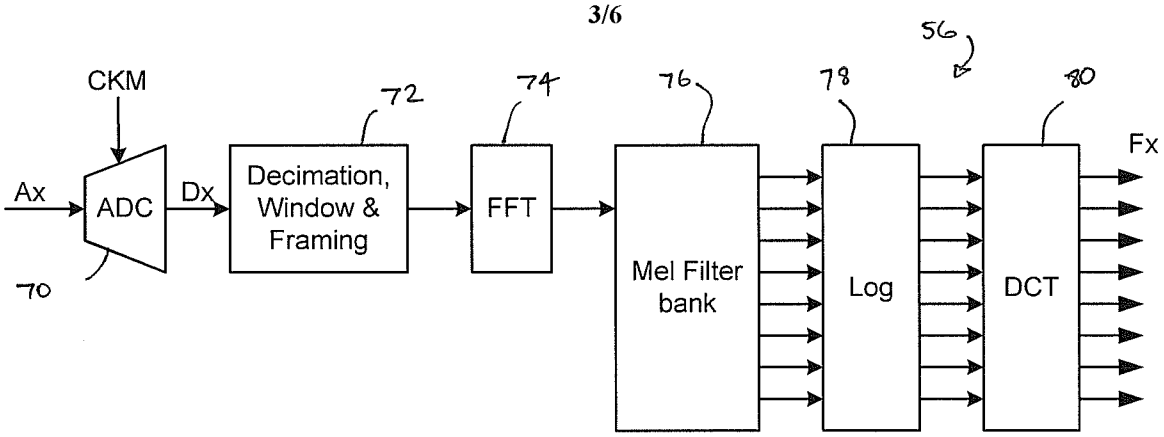


Figure 5

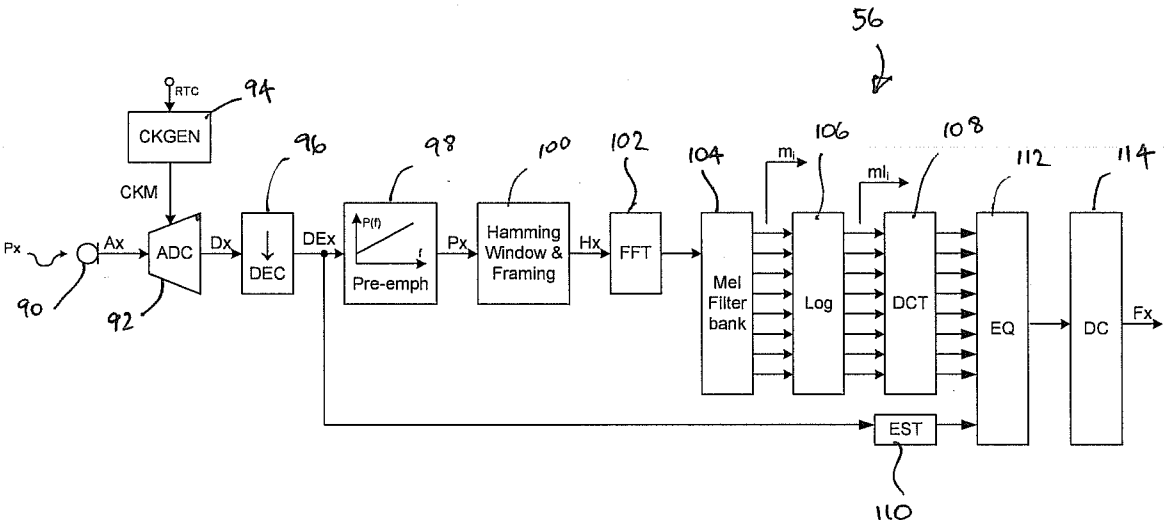


Figure 6

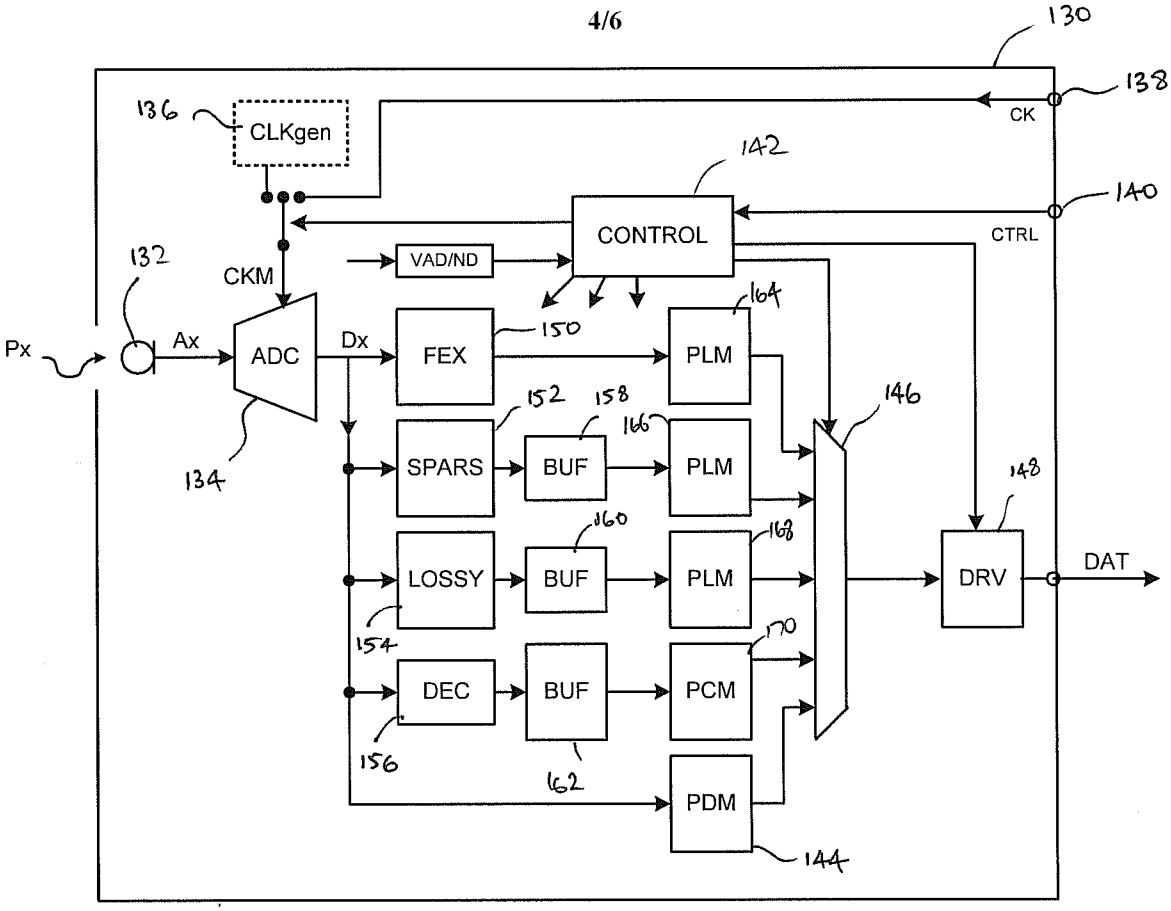

130
136
CLKgen
138
CK
142
CONTROL
140
CTRL
VAD/ND
132
CKM
Px
Ax
ADC
Dx
FEX
150
164
PLM
152
158
166
SPARS
BUF
PLM
146
148
160
168
LOSSY
BUF
PLM
DRV
DAT
154
170
DEC
BUF
PCM
156
162
PDM
144
134

Figure 7

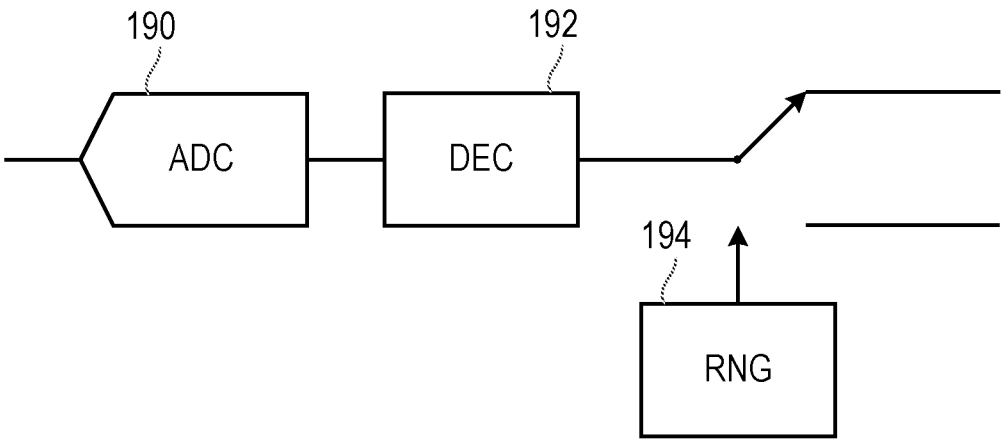


Figure 8

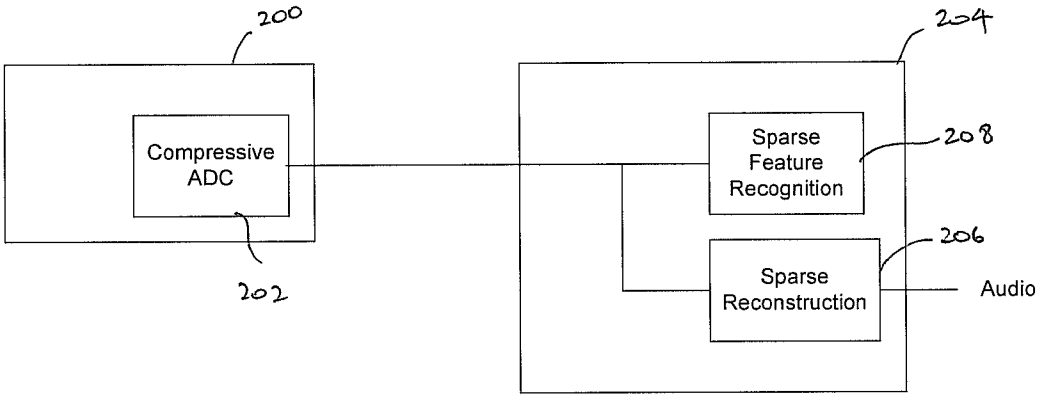


Figure 9

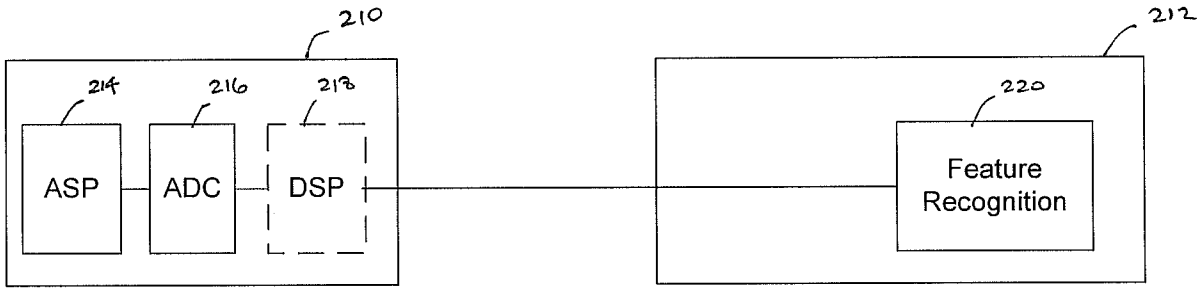


Figure 10

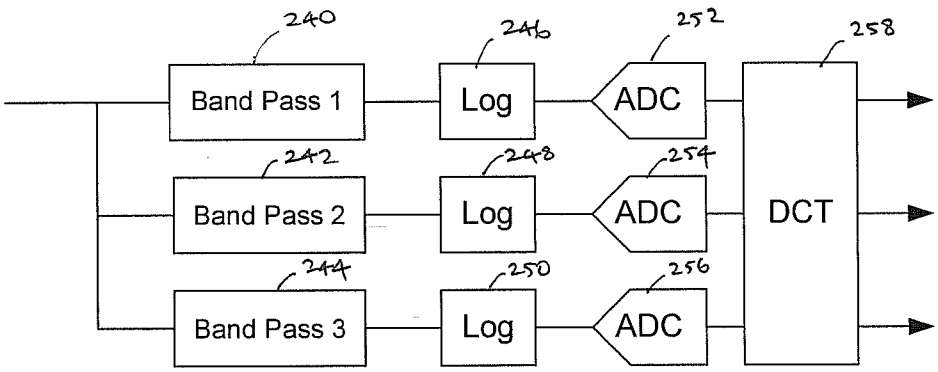


Figure 11

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2015/054122

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04R3/00 G10L25/24 G10L19/04
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L H04R

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2014/257813 A1 (MORTENSEN MIKAEL [DK]) 11 September 2014 (2014-09-11)	1-5, 9-14, 26-36, 40,56, 57,62-73
Y	paragraphs [0001] - [0008], [0024] - [0030], [0036] - [0043], [0050] - [0052] claims 1-5 figures 1-4	6,15-22, 41-50

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 March 2016	23/03/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Bensa, Julien

2

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2015/054122

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 612 974 A (ASTRACHAN PAUL M [US]) 18 March 1997 (1997-03-18) column 4, lines 13-28 column 5, line 46 - column 6, line 35 column 15, lines 1-33 figures 1,2	1,7,8, 26,28, 37-39, 56, 58-62,72
X	US 2006/034472 A1 (BAZARJANI SEYFOLLAH [US] ET AL) 16 February 2006 (2006-02-16) paragraphs [0005], [0020] - [0034]	1,23-26, 28, 51-56, 62,72
X	WO 2010/030889 A1 (PERSONICS HOLDINGS INC [US]; GOLDSTEIN STEVEN [US]; BOILLOT MARC [US];) 18 March 2010 (2010-03-18) figures 4-6 paragraphs [0059] - [0084]	1,26,28, 56,62,72
Y	JASON LASKA ET AL: "Random Sampling for Analog-to-Information Conversion of Wideband Signals", DESIGN, APPLICATIONS, INTEGRATION AND SOFTWARE, 2006 IEEE DALLAS/CAS W ORKSHOP ON, IEEE, PI, 1 October 2006 (2006-10-01), pages 119-122, XP031052636, ISBN: 978-1-4244-0669-2 the whole document	6
Y	EP 1 154 408 A2 (TOSHIBA KK [JP]) 14 November 2001 (2001-11-14) paragraphs [0014] - [0021], [0114]	15-22, 41-50
A	WO 2014/189931 A1 (KNOWLES ELECTRONICS LLC [US]) 27 November 2014 (2014-11-27) paragraph [0026]	21,22, 49,50

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2015/054122

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014257813 A1	11-09-2014	CN 104038864 A	10-09-2014
		US 2014257813 A1	11-09-2014
US 5612974 A	18-03-1997	NONE	
US 2006034472 A1	16-02-2006	CA 2576520 A1	23-02-2006
		EP 1782656 A1	09-05-2007
		JP 2008510378 A	03-04-2008
		US 2006034472 A1	16-02-2006
		WO 2006020478 A1	23-02-2006
WO 2010030889 A1	18-03-2010	EP 2332345 A1	15-06-2011
		JP 2012502596 A	26-01-2012
		US 2010142725 A1	10-06-2010
		US 2013279703 A1	24-10-2013
		WO 2010030889 A1	18-03-2010
EP 1154408 A2	14-11-2001	DE 60129072 T2	06-03-2008
		EP 1154408 A2	14-11-2001
		JP 2001318694 A	16-11-2001
		US 2001041976 A1	15-11-2001
		US 2005096904 A1	05-05-2005
WO 2014189931 A1	27-11-2014	CN 105379308 A	02-03-2016
		EP 3000241 A1	30-03-2016
		KR 20160010606 A	27-01-2016
		US 2014348345 A1	27-11-2014
		US 2015043755 A1	12-02-2015
		US 2015350774 A1	03-12-2015
		WO 2014189931 A1	27-11-2014