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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
01 February 2018 (01.02.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/021920 A1

(51) International Patent Classification:
A61B 5/087 (2006.01)

(21) International Application Number:
PCT/NZ2017/050104

(22) International Filing Date:
27 July 2017 (27.07.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
722578 27 July 2016 (27.07.2016) NZ

(71) Applicant: THE UNIVERSITY OF CANTERBURY [NZ/NZ]; Ilam Road, Private Bag 4800, Christchurch, 8140 (NZ).

(72) Inventors: DE RYBEL, Tom Gerard; Kantonstraat 47, 9320 Nieuwerkerken (BE). DERRICK, Donald James; 141 Homestead Lane, Ilam Apartments 141/02, Ilam, Christchurch, 8041 (NZ).

(74) Agent: AJ PARK; Level 22, State Insurance Tower, 1 Willis Street, Wellington (NZ).

(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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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).

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

(54) Title: MASKLESS SPEECH AIRFLOW MEASUREMENT SYSTEM

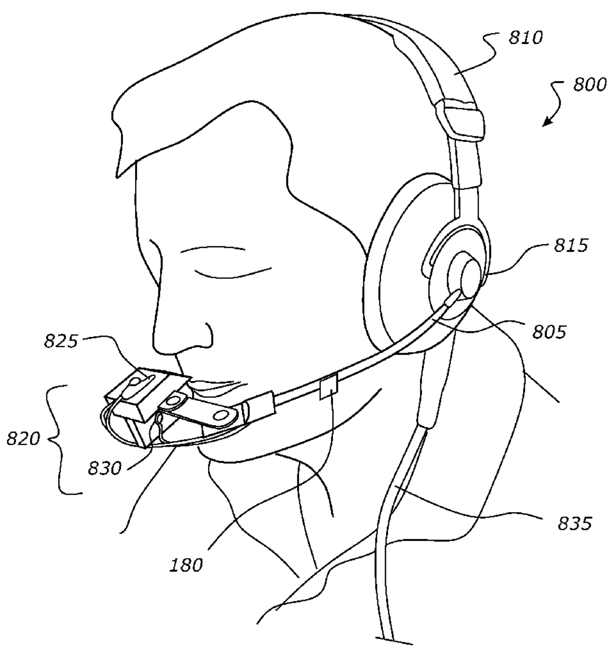

FIGURE 8

(57) Abstract: A speech airflow measurement system comprises a feature extraction module configured to receive input signals associated to a user from at least a first sensor and a second sensor, and determine an estimated shape and/or rate of airflow from at least part of the input signals. The system may further comprise a headset comprising the first sensor positioned within at least the first airflow of the user; the second sensor positioned within at least the second airflow of the user; and a shielding member adapted to shield the first sensor from the second airflow, the shielding member adapted to provide an air gap between the shielding member and a face of the user while the headset is in use by the user.

WO 2018/021920 A1

MASKLESS SPEECH AIRFLOW MEASUREMENT SYSTEM

FIELD OF THE INVENTION

This invention relates to a speech airflow measurement system.

BACKGROUND

Speech airflow measurement devices are commonly used by speech therapists to diagnose and treat speech disorders. The only way to measure airflow from speech precisely is to capture the air flow from the oral and nasal passages and pass it through a flow meter. To achieve accurate results, it is necessary to capture all the air flow and convey it through the flow meter.

Typical speech airflow recording techniques are based on pneumotachograph technology, which involves covering the mouth and/or nose with a mask and measuring air flow and air pressure differences across defined ports in the mask, or through connected tubes equipped with appropriate sensors. Other techniques involve either masks or sensory arrays that obscure the face of the speaker in order to collect speech airflow information.

While these systems provide detailed information about oral and nasal airflow, they obstruct the view of the speaker's face, and they may muffle speech so that the speaker is hard to understand. Wearing a mask, together with the distorted audio feedback to the speaker's ears, reduces the speaker's comfort and also impacts their mode of speaking.

Also, with a mask-based system, breathing often takes significantly more effort, due to the constriction of the air flow in the mask ports and/or sensors. This has the potential to affect speech production and breathing patterns, as well as induce a sense of claustrophobia or suffocation.

Visual obstruction is a further issue for speech research and some clinical applications. When people speak, they produce useful visual motion, flesh vibrations, sound and airflow. Sound travels as longitudinal waves through the air, whereas airflow from speech propagates air molecules out of the mouth and nose (and sometimes inward to the lungs). All of this information can be used to understand speech more effectively.

Finally, for many potential applications and social settings, such bulky masks are not acceptable, making several commercial uses of air flow information impractical with present technology.

One key, clinical application of a practical, maskless airflow estimation system is determining nasality. US patent specification 8,457,965 (Rothenberg) for example discloses a method of processing the ratio of nasal and oral airflow signals (nasalance) to correct for variation associated with different vowel sounds. Rothenberg also discloses a method for correcting recorded values of vowel nasalance by reducing the effect of crossover between the oral and nasal channels due to imperfect acoustic separation. Rothenberg's method is intended for use with conventional speech airflow measurement devices which require an acoustic separation means held against the upper lip of a user. Such devices suffer from the disadvantages described above.

It is an object of at least preferred embodiments of the present invention to provide a system for estimating speech airflow that preserves audio quality and feedback for a user.

It is a further object of at least preferred embodiments of the present invention to provide a system for estimating speech airflow that provides minimum discomfort and physical interference to the user.

It is a further object of at least preferred embodiments of the present invention to provide a system for estimating speech airflow that has minimal visual interference to allow for better non-verbal analysis and improve the social acceptance of the system.

An additional or alternative object of at least preferred embodiments of the present invention is to at least provide the public with a useful alternative.

SUMMARY OF THE INVENTION

In accordance with an aspect of the invention, a speech airflow measurement system comprises a feature extraction module configured to receive input signals associated to a user from at least a first sensor and a second sensor, and determine an estimated shape and/or rate of airflow from at least part of the input signals.

The term 'comprising' as used in this specification and claims means 'consisting at least in part of'. When interpreting statements in this specification and claims which include the term 'comprising', other features besides the features prefaced by this term in each statement can also be present. Related terms such as 'comprise' and 'comprised' are to be interpreted in a similar manner.

In an embodiment at least some of the input signals from the first sensor are associated to a first airflow of the user. The first airflow of the user comprises a nasal airflow.

In an embodiment at least some of the input signals from the second sensor are associated to a second airflow of the user. The second airflow of the user comprises an oral airflow.

In an embodiment the input signals from the first sensor and/or second sensor include input signals associated to the first airflow of the user and input signals associated to the second airflow of the user, the feature extraction module configured to separate the input signals associated to the first airflow of the user from the input signals associated to the second airflow.

In an embodiment the input signals from the first sensor and/or the second sensor are associated to at least one frequency in the range 0.1 Hz to 70 Hz. In an embodiment the input signals from the first sensor and/or the second sensor are associated to at least one frequency in the range 20 Hz to 20 kHz.

In an embodiment the input signals from the first sensor and/or the second sensor are associated to at least one time period in the range 14ms to 10s.

In an embodiment the feature extraction module further comprises an oral airflow estimator configured to estimate oral airflow and/or a nasal airflow estimator configured to estimate nasal airflow.

In an embodiment the feature extraction module further comprises an oral audio estimator configured to estimate oral audio and/or a nasal audio estimator configured to estimate nasal audio.

In an embodiment the feature extraction module further comprises a nasalance module configured to compute at least one nasalance indicator from the input signals from the first sensor and/or from the input signals from the second sensor.

In an embodiment the feature extraction module further comprises a fundamental frequency (f0) tracker configured to extract fundamental frequency from the input signals from the first sensor and/or the input signals from the second sensor.

In an embodiment the feature extraction module further comprises a breath group tracker configured to compute at least one breath group from the input signals from the first sensor and/or from the input signals from the second sensor.

In an embodiment the feature extraction module further comprises a formant tracker configured to extract at least one vowel formant from the input signals from the first sensor and/or from the input signals from the second sensor.

In an embodiment the feature extraction module further comprises a voice quality tracker configured to extract at least one voice quality indicator from the input signals from the first sensor and/or from the input signals from the second sensor.

In an embodiment the feature extraction module is configured to receive input signals associated to the user from a third sensor not associated to either the first airflow of the user or the second airflow of the user.

In an embodiment, the system further comprises a pre-processing module configured to apply at least one correction to the input signals from the first sensor and/or to the input signals from the second sensor; and generate an output for use by the feature extraction module.

In an embodiment, the system further comprises a classification module configured to receive as input at least part of the feature extraction module output; and generate diagnostic information as an output.

In an embodiment, the system further comprises a post-processing module configured to receive as input at least part of the feature extraction output and/or at least part of the classification module output; and generate a visual representation as an output.

In an embodiment the speech airflow estimation system further comprises a headset comprising the first sensor positioned within at least a first airflow of the user; the second sensor positioned within at least a second airflow of the user; and a shielding member adapted to shield the first sensor from the second airflow, the shielding member adapted to provide an air gap between the shielding member and a face of the user while the headset is in use by the user.

In an embodiment the headset further comprises a sensor assembly comprising a nasal sensor assembly and an oral sensor assembly, the nasal sensor assembly including the first sensor, the oral sensor assembly including the second sensor.

In an embodiment the nasal sensor assembly and/or the oral sensor assembly include(s) respective Venturi structures shaped to case a constriction of a flow path within the nasal sensor assembly and/or the oral sensor assembly.

In an embodiment, the headset further comprises the third sensor located substantially outside of the first airflow of the user and the second airflow of the user.

In an embodiment, at least one of the sensors comprises a sound pressure sensor. The at least one sensor may comprise a microphone, a strain gauge, or any other suitable sound pressure sensor.

In an embodiment, at least one of the sensors comprises an airflow velocity sensor. The at least one sensor may comprise a microphone, an anemometer, or any other suitable airflow velocity sensor.

In an embodiment, at least one of the sensors comprises an airflow sensor. At least one sensor may comprise a flow meter or an airflow velocity sensor in a constraining or partially constraining structure.

In an embodiment, the input signal is filtered so that the estimation of rate of airflow is based on 'pseudo-sound' signals. The input signal may be filtered using an inverse filter.

In an embodiment, the rate of airflow is estimated using input signals from the at least one sensor combined with an audio signal.

In accordance with a further aspect of the invention, a method of estimating speech airflow comprises receiving input signals associated to a user from at least a first sensor and a second sensor, and determining an estimated shape and/or rate of airflow from at least part of the input signals.

It is intended that reference to a range of numbers disclosed herein (for example, 1 to 10) also incorporates reference to all rational numbers within that range (for example, 1, 1.1, 2, 3, 3.9, 4, 5, 6, 6.5, 7, 8, 9 and 10) and also any range of rational numbers within that range (for example, 2 to 8, 1.5 to 5.5 and 3.1 to 4.7) and, therefore, all sub-ranges of all ranges expressly disclosed herein are hereby expressly disclosed. These are only examples of what is specifically intended and all possible combinations of numerical values between the lowest value and the highest value enumerated are to be considered to be expressly stated in this application in a similar manner.

This invention may also be said broadly to consist in the parts, elements and features referred to or indicated in the specification of the application, individually or collectively, and any or all combinations of any two or more said parts, elements or features, and where specific integers are mentioned herein which have known equivalents in the art to which this invention relates, such known equivalents are deemed to be incorporated herein as if individually set forth.

As used herein the term '(s)' following a noun means the plural and/or singular form of that noun.

As used herein the term 'and/or' means 'and' or 'or', or where the context allows both.

The invention consists in the foregoing and also envisages constructions of which the following gives examples only.

The terms ‘component', ‘module', ‘system', ‘interface', and/or the like as used in this specification in relation to a processor are generally intended to refer to a computer-related entity, either hardware, a combination of hardware and software, software, or software in execution. For example, a component may be, but is not limited to being, a process running on a processor, a processor, an object, an executable, a thread of execution, a program, and/or a computer. By way of illustration, both an application running on a controller and the controller can be a component. One or more components may reside within a process and/or thread of execution and a component may be localized on one computer and/or distributed between two or more computers.

The term 'connected to’ as used in this specification in relation to data or signal transfer includes all direct or indirect types of communication, including wired and wireless, via a cellular network, via a data bus, or any other computer structure. It is envisaged that they may be intervening elements between the connected integers. Variants such as 'in communication with', 'joined to', and 'attached to' are to be interpreted in a similar manner. Related terms such as 'connecting' and 'in connection with' are to be interpreted in the same manner.

In this specification where reference has been made to patent specifications, other external documents, or other sources of information, this is generally for the purpose of providing a context for discussing the features of the invention. Unless specifically stated otherwise, reference to such external documents or such sources of information is not to be construed as an admission that such documents or such sources of information, in any jurisdiction, are prior art or form part of the common general knowledge in the art.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described by way of example only and with reference to the accompanying drawings in which:

Figure 1 shows an example of a signal processing system adapted to receive and process speech airflow signals;

Figure 2 shows examples of modules implemented in the computing device of figure 1;

Figure 3 shows an example of the feature extraction module from figure 2;

Figure 4 shows an example of a computational technique for separating oral and nasal signals;

Figure 5 shows an example of audio and airflow estimation;

Figure 6 shows an example of nasalance;

Figure 7 shows an example of fundamental frequency (f0);

Figure 8 shows an example of a headset adapted for use with the speech airflow estimation system of figures 1 to 3; and

Figures 9 and 10 show the sensor assembly from figure 8 in more detail.

DETAILED DESCRIPTION

Preferred embodiments of the invention have been described by way of example only and modifications may be made thereto without departing from the scope of the invention.

For many speech airflow measurement applications it is sufficient to obtain an estimate of the shape and relative magnitude of the nasal and oral airflow, rather than a true measurement. Described below is a system for estimating speech airflow based on pressure and/or velocity measurements that does not require a mask and does not require a physical airflow separator that contacts the face of the user.

Speech airflow comprises two types of flow – laminar and turbulent. Laminar airflow moves in parallel layers with no interaction between the layers. Turbulent airflow involves chaotic property changes, with low momentum diffusion and high momentum convection. This means that turbulent airflow moves in vortices that mix the air up, and so moves away from the source more slowly than laminar airflow.

Sound is a longitudinal wave that propagates through air during speech. Sound produced by speech is usually a mixture of periodic and aperiodic longitudinal motion that ranges in frequency from about 20 Hz to about 20,000 Hz. Most of the sound energy from speech occurs in the frequency range of about 80 Hz to 8,000 Hz, depending on the age, body size and physical gender of the speaker.

Sound and airflow both produce changes in the surrounding air pressure and velocity that can be detected by sensors. If a pressure or velocity sensor is placed in front of a speaker, pressure or velocity changes from laminar airflow and turbulent airflow, as well as true sound, all interact with the sensor. It is very difficult to distinguish the respective contributions of changes due to sound and changes due to airflow from the sensor's response.

The airflow contribution to pressure and velocity changes detected by a sensor can be described as 'pseudo-sound'. Pseudo-sound is produced by the interaction of outgoing airflow from the mouth and nose interacting with the sensor. The laminar airflow will interact with the sensor in a relatively continuous fashion, and the turbulent airflow will cause random fluctuations in the sensor response. In contrast, true sound is a longitudinal wave that will create fluctuations in the sensor according to the frequency of the sound waves produced.

A key difference between speech airflow and speech sound is that speech airflow consists of a relatively large bulk volume of air molecules moving in, or out, of the mouth and nose. By contrast, speech sound is a longitudinal compression wave, which yields no net mass motion. The directionality and mass generated due to speech airflow means that speech airflow generates higher intensity activity at much lower frequencies than speech sound.

A common embodiment of an air pressure or velocity sensor is a microphone. Microphones typically comprise filters that supress low frequency information below about 20 Hz (known as 'high-pass' filtering). Their associated electronic circuits often incorporate filters to further suppress low frequencies. Low frequency information introduces artefacts into the recording, such as rubbing sounds and microphone 'pops', and reduces the amplitude difference that can be successfully recorded. Such artefacts can be introduced by a windy environment, an environment where an external object might touch or shake the microphone, or when a speaker breathes close to the microphone.

High-pass filtering is standard practice, as there is little useful speech audio below 20 Hz and human hearing has little sensitivity for such low frequencies.

The low frequency information that is usually filtered out of audio recordings has the potential to provide useful information about speech airflow. By using an unfiltered, or suitably filtered (with or without compensation), pressure or velocity sensor, speech airflow can be estimated.

This approach allows a speaker to be seen and heard clearly while airflow information is estimated. It also significantly reduces the bulk of the measurement system, making it suitable for non-research or clinical settings and applications, and situations in which social acceptance is important.

Figure 1 shows an example of a signal processing system adapted to receive and process speech airflow signals.

The operating environment of figure 1 is an example of a suitable operating environment. It is not intended to suggest any limitation as to the scope of use or functionality of the operating environment. Example computing devices include, but are not limited to, personal computers, server computers, hand-held or laptop devices, mobile devices, multiprocessor systems, consumer electronics, mini computers, mainframe computers, and distributed computing environments that include any of the above systems or devices. Examples of mobile devices include mobile phones, tablets, and Personal Digital Assistants (PDAs).

Although not required, embodiments are described in the general context of 'computer readable instructions' being executed by one or more computing devices. In an embodiment, computer readable instructions are distributed via tangible computer readable media.

In an embodiment, computer readable instructions are implemented as program modules. Examples of program modules include functions, objects, Application Programming Interfaces (APIs), and data structures that perform particular tasks or implement particular abstract data types. Typically, the functionality of the computer readable instructions is combined or distributed as desired in various environments.

Shown in figure 1 is a system 100 comprising a computing device 105 configured to implement one or more embodiments described below. In an embodiment, computing device 105 includes at least one processing unit 110 and memory 115. Depending on the exact configuration and type of computing device, memory 115 is volatile (such as RAM, for example), non-volatile (such as ROM, flash memory, etc., for example) or some combination of the two.

A server 120 is shown by a dashed line notionally grouping processing unit 110 and memory 115 together.

In an embodiment, computing device 105 includes additional features and/or functionality.

One example is removable and/or non-removable additional storage including, but not limited to, magnetic storage and optical storage. Such additional storage is illustrated in figure 1 as storage 125. In an embodiment, computer readable instructions to implement one or more embodiments provided herein are maintained in storage 125. In an embodiment, storage 125 stores other computer readable instructions to implement an operating system and/or an application program. Computer readable instructions are loaded into memory 115 for execution by processing unit 110, for example.

Memory 115 and storage 125 are examples of computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, Digital Versatile Disks (DVDs) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by computing device 105. Any such computer storage media may be part of device 105.

In an embodiment, computing device 105 includes at least one communication connection 140 that allows device 105 to communicate with other devices. The at least one communication connection 140 includes one or more of a modem, a Network Interface Card (NIC), an integrated network interface, a radio frequency transmitter/receiver, an infrared port, a USB connection, or other interfaces for connecting computing device 105 to other computing devices.

In an embodiment, the communication connection(s) 140 facilitate a wired connection, a wireless connection, or a combination of wired and wireless connections. Communication connection(s) 140 transmit and/or receive communication media.

Communication media typically embodies computer readable instructions or other data in a "modulated data signal" such as a carrier wave or other transport mechanism and includes any information delivery media. The term "modulated data signal" includes a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal.

In an embodiment, device 105 includes at least one input device 145 such as keyboard, mouse, pen, voice input device, touch input device, infrared cameras, video input devices, and/or any other input device. In an embodiment, device 105 includes at least one output device 150 such as one or more displays, speakers, printers, and/or any other output device.

Input device(s) 145 and output device(s) 150 are connected to device 105 via a wired connection, wireless connection, or any combination thereof. In an embodiment, an input device or an output device from another computing device is/are used as input device(s) 145 or output device(s) 150 for computing device 105.

In an embodiment, components of computing device 105 are connected by various interconnects, such as a bus. Such interconnects include one or more of a Peripheral

Component Interconnect (PCI), such as PCI Express, a Universal Serial Bus (USB), firewire (IEEE 13104), and an optical bus structure. In an embodiment, components of computing device 105 are interconnected by a network. For example, memory 115 in an embodiment comprises multiple physical memory units located in different physical locations interconnected by a network.

It will be appreciated that storage devices used to store computer readable instructions may be distributed across a network. For example, in an embodiment, a computing device 155 accessible via a network 160 stores computer readable instructions to implement one or more embodiments provided herein. Computing device 105 accesses computing device 155 in an embodiment and downloads a part or all of the computer readable instructions for execution. Alternatively, computing device 105 downloads portions of the computer readable instructions, as needed. In an embodiment, some instructions are executed at computing device 105 and some at computing device 155.

A client application 165 enables a user experience and user interface. In an embodiment, the client application 165 is provided as a thin client application configured to run within a web browser. The client application 165 is shown in figure 1 associated to computing device 155. It will be appreciated that application 165 in an embodiment is associated to computing device 105 or another computing device.

Shown in figure 1 is a first sensor 170 and a second sensor 175. The sensors are connected to computing device 105 via a wired connection, wireless connection, or any combination thereof. In an embodiment the first sensor 170 is positioned within at least a first airflow of a user. One example of a first airflow is nasal airflow. The signals generated by the first sensor 170 are associated to a nasal airflow of a user.

In an embodiment the second sensor 175 is positioned within at least a second airflow of a user. One example of a second airflow is oral airflow. The signals generated by the second sensor 175 are associated to an oral airflow of a user.

In an embodiment the first sensor 170 and the second sensor 175 each comprise respective microphones configured to detect airflow pressure of a sound in a pseudo-sound frequency range of about 0.1 to 70 Hz. As used herein the term 'low frequency' or 'pseudo-sound' denotes this frequency range of about 0.1 to 70 Hz.\In an embodiment the first sensor 170 and the second sensor 175 each comprise respective microphones configured to detect airflow pressure of a sound in a frequency range of about 20 Hz to about 20 kHz. This is in what is generally considered the audible range.

Acoustic microphones have the potential to be suitable sensors for capturing nasal and oral airflow information. Microphones are mass-produced and compact. However, the low frequency response of most commercial microphones is insufficient as they typically have built in high-pass filters. This shortcoming can be improved with the use of an inverse filter, designed to compensate to a useful extent for the loss of microphone sensitivity and associated phase shift due to the acoustic high-pass filter in the microphone and/or the microphone pre-amplifier.

Similarly, the use of a suitable signal path in the capture device, implementing suitable high-pass filters or avoiding them altogether, is preferred to maintain the air flow information in the signal.

Similarly, the use of microphones without a high-pass filter, or a suitable high-pass filter, is preferred to maintain the air flow information in the signal.

A suitable high-pass filter is one so arranged as to pass information above about 0.1 Hz, while rejecting lower frequency (very slow) pressure variations, such as those caused by atmospheric pressure variations, static pressure from building ventilation systems, changes in altitude, and so on, that would create large offsets in the microphone output unrelated to the airflow information.

In an embodiment, the high-pass filter comprises a vented chamber in the device, one side of it the microphone membrane, where the venting port in this chamber to the outside sets the high-pass cut-off frequency. In an embodiment the high-pass filter forms part of the pressure equalization system.

Suitable types of acoustic microphone include MEMs, electret microphones (self-polarized, capacitive), condenser microphones (externally-polarized, capacitive), other electret/condenser microphones (for example Sennheiser's modulated bridge design), piezo-electric microphones, dynamic (moving-coil) inductive microphones, optical microphones, and interferometer microphones.

In alternative embodiments, other types of sound pressure sensor are used, for example strain gauges of the resistive type (in all common configurations, such as full bridge, single element, half-bridge, and all their modes of excitation), and strain gauges of the capacitive type (in all their common configurations, and all their modes of excitation).

In alternative embodiments, other types of airflow velocity sensor are used, for example thermal sensors such as hot wire anemometers, ultrasonic (acoustic) flow meters, optical flow meters based on transmitted or reflected light from particles in the air flow (eg: water droplets), interferometers, and mechanical sensors, such as (rotating) vanes, oscillating structures, vortex shedding, and in-flow deflection sensors. Other types are based on measuring pressure differences across a constriction in the flow path, such as a venturi, or an acoustic impedance in the flow path, such as a fine wire mesh.

The computing device 105 is adapted to receive signals from the first sensor 170 and/or second sensor 175. The computing device is adapted to process the signals and generate output for an output device 150 for example a display.

Figure 2 shows examples of modules implemented in one embodiment of the computing device 105. It will be appreciated that in some embodiments not all modules shown in figure 2 are implemented on computing device 105. It will also be appreciated that in some embodiments additional modules are implemented on computing device 105.

Shown in figure 2 is a pre-processing module 200 adapted to apply corrections to the signals from the first sensor 170 and/or second sensor 175. As described above, the first sensor 170 is associated to nasal airflow and the second sensor 175 is associated to oral airflow. In an embodiment the pre-processing module 200 is further configured to generate an output for use by the feature extraction module 205.

In an embodiment, the sensor signal from the sensors 170 and 175 is pre-processed in the pre-processing module 200 and at least one transfer function of the sensors and their interface corrected to achieve a wider useful frequency response. In an embodiment the pre-processing module 200 is adapted to correct for other short-comings of the sensor and/or signal path, including but not limited to: recovering from sensor overload conditions due to plosives (short but intense bursts of air during eg 'p' sounds), DC baseline drift, and time-aligning the signals to compensate for potentially unequal filter delay as well as energy propagation delays between the mouth/nose and oral/nasal sensors. These delays may be frequency-dependent.

In an embodiment the pre-processing module 200 is adapted to correct for undesirable cross-coupling between the oral and nasal signals in the sensor signals. To minimise this undesirable cross-coupling, good placement of the sensors, and suitable physical structures (such as the shielding structures) positively impact the degree of separation between oral and nasal information in the sensor signals.

In an embodiment shown in figure 2, the pre-processing module 200 uses data from the first sensor 170 to filter data from the data recorded by the second sensor 175. The pre-processing module 200 in an embodiment uses data from the second sensor 175 to filter data from the data recorded by the first sensor 170. Any combination of using data from at least one sensor to filter data from at least one other sensor is possible in various embodiments. For example, in an embodiment the third sensor 180 is configured to reduce background noise interference.

In an embodiment a feature extraction module 205 is configured to receive input signals from sensor 170 and/or sensor 175. In an embodiment a feature extraction module 205 is configured to receive pre-processed input signals from sensor 170 and/or sensor 175 via pre-processing module 200. In an embodiment the feature extraction module is configured to receive input signals directly from sensor 170 and/or sensor 175. As described above, the first sensor 170 is associated to nasal airflow and the second sensor 175 is associated to oral airflow.

In an embodiment a third sensor 180 is positioned outside the nasal airflow captured by the first sensor 170 and outside the oral airflow captured by the second sensor 175. The third sensor 180 is configured to obtain an audio signal that is substantially free from any artefacts due to any pseudo sound impacting the first sensor 170 and/or second sensor 175.

The third sensor 180 has the potential to be used for noise reduction purposes in circumstances where the system is used in less than ideal conditions.

The feature extraction module 205 is further configured to calculate estimated rate of airflow from at least part of the signals from the sensor(s) 170 and 175 using one or more mathematical methods.

In an embodiment, the airflow is estimated from the frequency range in which most of the non-sound airflow energy impacts the sensors. In an embodiment the frequency from which airflow is estimated is in the range 0.1 Hz to 70 Hz. The estimation is made within a time period of 14 ms to 10 s.

An embodiment of the speech airflow estimation technique takes as input the force from the airflow applied to a pressure or velocity or flow transducer. This provides a reasonable approximation of airflow change over time that relates well in time and space to individual speech sounds, or phones, as well as words, phrases, or speech breath cycles.

Shown in figure 2 is a classification module 210. In an embodiment the classification module 210 is adapted to take input from the feature extraction module 205 and is adapted to output diagnostic information. Input obtained from the feature extraction module 205 includes nasalance (a measure of the nasality of speech), fundamental frequency (f0) tracking (a measure corresponding to perceived pitch), and vowel formant tracking (a measure of vowel quality).

In an embodiment, the classification module 210 is adapted to identify nasal airflow in speech and/or oral airflow in speech.

In an embodiment, the classification module 210 is adapted to capture physical speech impairment symptoms and/or chemical speech impairment symptoms. This includes but is not limited to, hyper or hypo-nasality, incorrect oral and nasal airflow patterns, incorrect prosody (as measured in part by fundamental frequency tracking, and in part by speech timing), incorrect tone (as measured by the fundamental frequency tracking), and incorrect pronounciation or word choice (as captured through the audio portion of the signal.)

In an embodiment, the classification module 210 is used to identify one or more of speech breathing irregularities, phonation irregularities, speech production irregularities, velopharyngeal inadequacy from any cause, lip closure inadequacy from any cause, language and accent.

Shown in figure 2 is a post-processing module 215 adapted to store the processed data. In an embodiment the post-processing module 215 is adapted to receive input data from the classification module 210 and/or from the feature extraction module 205. The post-processing module 215 is adapted to receive input data directly from the feature extraction module 205 in embodiments that do not include a classification module 210.

In an embodiment, the post-processing module 215 stores, in at least one data file, one or more of: raw sensor information, audio, nasal airflow, oral airflow, and/or computed quantities from this information such as nasalence, fundamental frequency or any combination of the foregoing.

In an embodiment, the post-processing module 215 generates a visualisation or visual representation of the raw sensor information, audio, nasal airflow, and/or oral airflow output, and/or computed quantities from this information such as nasalence, fundamental frequency, or any combination of the foregoing, in a processed visualization and/or a real-time visualization. The visualisations are displayed for example on an output device 150 from figure 1 such as a display.

In an embodiment, the post-processing module 215 is adapted to display one or more of identifying markers for physical speech impairment symptoms, identifying markers for chemical speech impairment symptoms, identifying markers for speech disability diagnosis, including but not limited to speech breathing issues, phonation issues, or acute speech production issues, identifying markers for velopharyngeal inadequacy from any cause, identifying markers for lip closure inadequacy from any cause, identifying markers for identifying language and accent.

In an embodiment, the post-processing module 215 is adapted to provide a biofeedback tool for purposes including but not limited to, language training, accent training, or training for people with speech and/or hearing disabilities.

Figure 3 shows an example embodiment of the feature extraction module 205 from figure 2.

In an embodiment, the feature extraction module 205 includes an oral airflow estimator 300 configured to estimate oral airflow and/or a nasal airflow estimator 305 configured to estimate nasal airflow.

In an embodiment, the oral flow estimator 300 and/or nasal flow estimator 305 includes at least one filter configured to estimate airflow. Examples of filters include low-pass filters and band-pass filters.

The oral flow estimator 300 and/or nasal flow estimator 305 are configured in various embodiments to estimate airflow based at least partly on one or more of: zero-crossing rates, periodicity, an autocorrelation, an instantaneous frequency, a frequency energy, a statistical measure, a rate of change, an intensity root mean square value, time spectral information, a filter, a filter bank, a demodulation scheme, time window gating, formant tracking, or the acoustic signal itself.

In an embodiment, the oral flow estimator 300 and/or nasal flow estimator 305 are configured to estimate airflow based at least partly on one or more of: heuristics, logic systems, mathematical analysis, statistical analysis, learning systems, gating operation, range limitation, and normalization on the candidate air flow portion.

In an embodiment, the rate of airflow is estimated using input signals from the sensor 170 and/or sensor 175 combined with an audio signal. In an embodiment, airflow extraction is accomplished by filtering out the audio signal, leaving the nasal and/or oral airflow. An oral audio estimator 310 is configured to obtain oral audio from the oral signal. A nasal audio estimator 315 is configured to obtain nasal audio from the nasal signal.

In an embodiment, the oral and/or nasal audio is filtered so that the audio portion is based on a reasonable acoustic range perceived by humans. In an embodiment, the oral audio estimator 310 and/or nasal audio estimator 315 includes at least one filter configured to estimate airflow. Examples of filters include high-pass filters and band-pass filters.

In an embodiment, one or more of the estimation filters (oral flow estimator 300, the nasal flow estimator 305, the oral audio estimator 310 and the nasal audio estimator

315) is configured so that their group delays (frequency-dependant time delay) are corrected for and the time-domain signals obtained after filtering for airflow and/or audio are time-aligned.

In an embodiment, the estimation filters are complemented with group delay (frequency-dependent time delay) compensation filters so the time-domain signals obtained after filtering for airflow and/or audio are time-aligned.

In an embodiment, a fundamental frequency (f0) tracker 320 is configured to extract fundamental frequency from the signal. The fundamental frequency is related to pitch perception. F0 variations contribute to prosody and in tonal languages help distinguish lexical categories.

The f0 tracker has the potential to provide the system with uses including language accentedness training and learning of tonal languages.

In an embodiment, a voice quality tracker 325 is configured to extract at least one voice quality indicator from the signal. The voice quality tracker 325 is configured to apply one or more of harmonic spacing, harmonic ratios, jitter, shimmer, harmonicity.

In an embodiment, a formant tracker 330 is configured to extract at least one vowel formant from the signal. The formant tracker 330 is configured to implement linear predictive coding.

In an embodiment, a breath group tracker 335 is configured to extract at least one breath group from the signal. The breath group tracker 335 is configured to implement sound/silence tracking.

In an embodiment, a nasalance module 340 is configured to compute at least one nasalance indicator from the oral and nasal signal channels. In an embodiment, nasalance is computed at least partly from amplitude of the nasal signal divided by the amplitude of the oral signal plus the amplitude of the nasal signal. In an embodiment, nasalance is computed using a nasalance formula filtered to exclude extraneous data such as airflow.

The nasalance module has the potential to provide the system with clinical uses for cleft palate and deaf speech, as well as accentedness training.

Figure 4 shows an example of a computational method for separating oral and nasal signals. In an embodiment, the method comprises an adaptive filter-based technique.

In an embodiment the source separation method at least partially undoes the mixing of the oral and nasal signals at each sensor by estimating the mixing coefficients. The mixing coefficients can be simplified to constants, or treated as frequency-dependent.

The method further comprises at least partially undoing the linear process of summation that occurs at each sensor. The flow and acoustic separation is never perfect in practice, therefore some correction may be required or at least advantageous.

In an embodiment, the coefficients for the methods are obtained by a calibration technique, where the user follows a specific procedure and generates oral and/or nasal cues accordingly.

In an embodiment, varying degrees of blind source separation are employed. The method, using for instance cross-correlations between the signals, automatically tunes adaptive filters to remove the contribution of one flow to the other using optimization techniques. One example of an optimisation technique is a gradient descent algorithm such as least-squares.

It is advantageous to control the filter adaptation process using a voice activity detection scheme, so that the adaptation only proceeds during those portions of speech where useful signals are present.

A goal of the method 400 shown in figure 4 is to measure the real oral and/or nasal signals associated for example to sensors 170 and 175 respectively.

However, due to the physics of the measurement system, part of the signal from 405 is summed to the signal 410, yielding an actually measured signal 415. The signal 405 is modified by its passage through the air and mechanical structures, a process which is represented by an unknown filter 420 and an unknown delay 425.

Conversely, the signal part of the signal from 410 is summed to the signal 405, yielding the actually measured signal 430. The signal 410 is modified by its passage through the air and mechanical structures, a process which is represented by an unknown filter 435 and an unknown delay 440.

Assuming these are sufficiently linear processes, inverse filters are implemented that undo the summation process. With appropriate inverse filters, estimates of the real oral and nasal signals are obtained through subtracting inverse-filtered versions of the mixed signals. In the embodiment shown, adaptive filters 455 and 460 are used.

Thus, signal 430, minus the estimated contribution of 410, becomes signal 445, which is an estimate of 405. Conversely, signal 415, minus the estimated contribution of 405, becomes signal 450, which is an estimate of 410.

Due to the cross-coupling of the signals, this process exhibits an iterative convergence as adaptive filters 455 and 460 adapt to minimize the cross-coupling.

The inverse filters 455 and 460 are obtained for example by minimizing cross-correlations through a gradient descent technique, such as least-squares. As the oral and nasal signals have quite different wave shapes, their natural cross-correlation is below unity, and a degree of separation using such a method is possible.

Figures 5 to 7 show examples of biofeedback displayed for example on the output device 150 of figure 1.

Figure 5 shows an example of audio and airflow estimation from an ideal speech example shown at 500. Airflow for the estimation is shown at 505. An audio waveform is shown for example at 510.

An example of speech obtained from a user is shown at 520. Airflow for the speech sample is shown at 525. An audio waveform is shown for example at 530.

Figure 6 shows an example of nasalance. Nasalance from an ideal speech example is shown at 600. Nasalance recorded from a user is shown at 605.

Figure 7 shows an example of fundamental frequency (f0). Fundamental frequency from an ideal speech example is shown at 700. Fundamental frequency recorded from a user is shown at 705.

Figure 8 shows an example of a headset 800 adapted for use with the speech airflow estimation system described above.

Headset 800 comprises a boom-arm 805 rotatably connected to a headband 810 via hinge 815. Also shown adjacent the hinge 815 is a pad configured to enclose an ear of the user. It will be appreciated that this pad is an optional feature of the headset 800.

A sensor assembly 820 is mounted at an end of the boom-arm 805 distal from the hinge 815. In an embodiment, the sensor assembly 820 comprises a nasal sensor assembly 825 and an oral sensor assembly 830. When the headset 800 is fitted to the head of a user, the nasal sensor assembly 825 is elevated with respect to the oral sensor assembly 830.

The sensor assembly 820 includes the first sensor 170 mounted within the nasal sensor assembly 825 at least partially within a nasal flow of the user. The sensor assembly 820 further includes the second sensor 175 mounted within the oral sensor assembly 830 at least partially within an oral flow of the user.

In an embodiment, a communication cable 835 connects the headset 800 to other components of the airflow estimation system, for example computing device 105. In an embodiment the connection between the headset 800 and computing device 105 is wireless thereby obviating the need for a communication cable 835.

In an embodiment, the communication cable 835 includes the sensor interface electronics, analog to digital converters, power provisions and interface circuits required to connect to a computing device 105 through wired or wireless means.

In an embodiment the headset 800 further comprises the third sensor 180 adapted to record audio. One example of a third sensor is a microphone. In an embodiment the third sensor 180 is located on the boom arm 805 substantially outside of the sources of oral and nasal airflow. In an embodiment the third sensor 805 is located substantially within the source of oral and/or nasal flow, for example within the sensor assembly 820.

In an embodiment, the third sensor 180 is contained within a shielding device adapted to shield the third sensor 180 from low frequency signals. Alternatively or additionally the output from the third sensor 180 is filtered to suppress low frequency signals.

Figure 9 and **Figure 10** show different views of an embodiment of the sensor assembly 820 in more detail.

As described above, the sensor assembly 820 comprises a nasal sensor assembly 825 and an oral sensor assembly 830. Nasal sensor assembly 825 includes first sensor 170. Oral sensor assembly 830 includes second sensor 175.

In an embodiment the sensor assembly 820 includes a shielding member 900. The shielding member 900 is shaped to include a first planar surface and a second planar surface. The nasal sensor assembly 825 is mounted to the first planar surface. The oral sensor assembly 830 is mounted to the second planar surface.

The shielding member 900 is shaped and positioned so as to provide a partial separation between a nasal airflow of the user and an oral airflow of the user. The shielding member 900 is shaped and positioned so as to shield the first sensor 170 within the nasal sensor assembly 825 and/or the second sensor 175 within the oral sensor assembly 830 from undesired airflows.

The shielding member 900 is further shaped and positioned so as to provide a mount to attach the sensor assembly 820 to the boom arm 805.

In an embodiment, the nasal sensor assembly 825 includes a pair of side walls 905 and 910 defining a cavity between side walls 905 and 910. The cavity extends through the nasal sensor assembly 825. In an embodiment, the width of the cavity is non-uniform along a length of the cavity. A pair of extensions, one of which is indicated at 915, extend from side walls 905 and 910 respectively to form a Venturi-like structure to which the first sensor 170 is connected through an aperture 925.

In an embodiment, the oral sensor assembly 830 includes a Venturi-like structure similar to that of the nasal sensor assembly 825. The oral sensor assembly 830 includes a pair of side walls 930 and 935 defining a cavity between side walls 930 and 935. The cavity extends through the oral sensor assembly 830. In an embodiment, the width of the cavity is non-uniform along a length of the cavity. A pair of extensions 940 and 945 extend within the cavity to form a Venturi-like structure to which the second sensor 175 is connected through an aperture 950.

The Venturi structures of the nasal sensor assembly 825 and/or the oral sensor assembly 830 are configured to at least partially reduce the air pressure at the sensors 170 and 175 to better match the dynamic range of the microphone to the high pressure events due to (pseudo) speech airflow, funnel the air flow from the nose and mouth to the sensors 170 and 175, and/or generate a degree of turbulence that makes flow more easily identified and audible.

The Venturi structures are shaped to cause a constriction of the flow path within the nasal sensor assembly 825 and/or the oral sensor assembly 830, which causes the pressure to decrease, but the flow velocity to increase. The Venturi structures reduce the pressure of the airflow locally, to better match the pressure of this flow to the dynamic range of the sensor.

The Venturi structures also serve to constrain and duct part of the flow, provide separation between the oral and nasal flows, and ensure a more turbulent nature of the flow, which is easier to capture using some types of sensors, such as microphones.

In the embodiments shown in figures 9 and 10, the Venturi structures have only one explicit sensing port, in the constricted part of the Venturi structure. The second port is the environment, which is shared by the sensor reference port, for example the breathing port in an acoustic microphone to the environment.

Two Venturi structures are used to arrive at a convenient mechanical assembly incorporating the sensor elements.

In an embodiment, at least one shielding device is adapted to perform frequency-selective shielding. In an embodiment, the frequency-selective shielding allows low frequency signals and attenuates high frequency signals. In an embodiment, the frequency-selective shielding allows high frequency signals and attenuates low frequency signals.

In an embodiment, the aperture 925 is formed in the nasal sensor assembly 825 at a location within the assembly 825 corresponding to the narrowest width of the cavity of the assembly 825. Similarly, the aperture 950 is formed in the oral sensor assembly 830 at a location within the assembly 830 corresponding to the narrowest width of the cavity of the assembly 830. The sensors 170 and 175 are positioned so as to measure airflow though the apertures 925 and 950 respectively.

In an embodiment, the width of the nasal assembly side wall 905 and nasal assembly side wall 910 is greater at a lower portion of the nasal sensor assembly 820 than at an upper portion. This difference in relative lengths, combined with the shielding member 900, serves to capture nasal airflow more effectively. Nasal flow enters largely from above into the nasal sensor assembly 825. The shielding member 900 at least partially channels the nasal flow into the nasal sensor assembly 825.

The nasal sensor assembly 825 and the oral sensor assembly 830 are positioned relative to each other so that the separation of oral and nasal flow paths is improved, and also the propagation path of the oral and nasal flows to the respective sensor structures is made more equal.

In an embodiment, the oral sensor assembly 830 is set back from the face in relation to the nasal sensor assembly 825. The lips of a user tend to protrude from the face meaning that a set back is desirable. Another factor is that pursed lips should preferably not touch any part of the sensor assembly 820 for comfort, but also to reduce rubbing noises.

In an embodiment, the nasal sensor assembly 825 is shaped with a width such that most of the airflow from the nostrils is captured. The oral sensor assembly 830 is shaped with a height such that a sufficient portion of the airflow from the mouth, even when lip positions shift during speech production.

In an embodiment, these relative dimensions are also somewhat over-sized to help accommodate differences in the relative placement of the sensor assembly 820 with respect to the user's face, as well as various facial geometries.

In an embodiment the sensor assembly 820 is positioned so that it does not touch the face of the user at rest and during speech. This has the potential to improve the comfort for the user, prevent unnatural speech patterns, and prevent structure-born interference noises in the sensor data due to rubbing and impacting of the sensor assembly with the face and lips during use.

In an embodiment the nasal sensor assembly 825 and oral sensor assembly 830 form a duct which further shields their respective sensors 170 and 175 from spurious airflow, as well as provide some degree of uniformity in the airflow that is captured during speech, by offering a defined orifice that funnels a defined portion of the respective flows to the sensors 170 and 175.

It will be appreciated that sensor assembly 820 comprising a nasal sensor assembly 825 and oral sensor assembly 830 is one example of an assembly. In an embodiment a sensor assembly includes a first sensor, a second sensor, and optionally a small separator/shielding structure. The first sensor and second sensor are positioned relative to the nose and lips of a user to achieve some degree of initial separation.

Energy transfer through the physical structures of the sensor assembly 820, as well as through the air is unavoidable. The use of mathematical source separation techniques such as those described above have the potential to enhance the degree of oral and nasal energy separation between the first sensor 170 and second sensor 175 beyond that which is obtainable through a physical separation alone.

The system described above has the potential to demonstrate nasality to a user in real-time. The user is able to use this biofeedback to change the amount of nasality in the user's speech.

One application for nasality biofeedback is accent modification. By making the user aware of the oral and nasal production patterns, speakers can be trained to produce the desired patterns of oral and nasal airflow in speech.

Another application for nasality biofeedback is voice training (singing lessons), where the user can gain information about their detailed speech (song) airflow patterns, in addition to the existing audio, and chest/midriff/belly movement (large-scale breathing patterns) capturing systems.

Another application is treatment for speech impediments due to one or more of traumatic brain injury, deaf speech, cleft lip/palate speech, or any other issues that influence oral airflow, nasal airflow, or nasality in speech.

Another application is telemedicine. Visualisations of oral and nasal production are transmitted to a therapist who is physically distant from the patient. The therapist can thus interact with a remote patient, simultaneously receiving audio, video, and oral/nasal flow information. This information enables various forms of therapy, that at present require the therapist to be present to feel the oral/nasal production, to be offered to patients who would otherwise have no access to such therapy (elderly, disabled, remote areas).

By incorporating bio-feedback principles and training tools a user is able to practice at home. Bio-feedback includes showing the flow information, a processed form of the flow information (such as a nasalence measure), computed features (such as f0), or indirect visualisations and game objectives.

Preferred embodiments of the invention have been described by way of example only and modifications may be made thereto without departing from the scope of the invention, as defined by the accompanying claims.

CLAIMS

1. A speech airflow measurement system comprising:

a feature extraction module configured to receive input signals associated to a user from at least a first sensor and a second sensor, and determine an estimated shape and/or rate of airflow from at least part of the input signals.

2. The system of claim 1 wherein at least some of the input signals from the first sensor are associated to a first airflow of the user.

3. The system of claim 2 wherein the first airflow of the user comprises a nasal airflow.

4. The system of claim 2 or 3 wherein at least some of the input signals from the second sensor are associated to a second airflow of the user.

5. The system of claim 4 wherein the second airflow of the user comprises an oral airflow.

6. The system of claim 4 or 5 wherein the input signals from the first sensor and/or second sensor include input signals associated to the first airflow of the user and input signals associated to the second airflow of the user, the feature extraction module configured to separate the input signals associated to the first airflow of the user from the input signals associated to the second airflow.

7. The system of any one of the preceding claims wherein the input signals from the first sensor and/or the second sensor are associated to at least one frequency in the range 0.1 Hz to 70 Hz.

8. The system of any one of the preceding claims wherein the input signals from the first sensor and/or the second sensor are associated to at least one frequency in the range 20 Hz to 20 kHz.

9. The system of any one of the preceding claims wherein the input signals from the first sensor and/or the second sensor are associated to at least one time period in the range 14ms to 10s.

10. The system of any one of the preceding claims wherein the feature extraction module further comprises an oral airflow estimator configured to estimate oral airflow and/or a nasal airflow estimator configured to estimate nasal airflow.

11. The system of any one of the preceding claims wherein the feature extraction module further comprises an oral audio estimator configured to estimate oral audio and/or a nasal audio estimator configured to estimate nasal audio.

12. The system of any one of the preceding claims wherein the feature extraction module comprises a nasalance module configured to compute at least one nasalance indicator from the input signals from the first sensor and/or from the input signals from the second sensor.

13. The system of any one of the preceding claims wherein the feature extraction module comprises a fundamental frequency (f0) tracker configured to extract fundamental frequency from the input signals from the first sensor and/or the input signals from the second sensor.

14. The system of claim 4 or 5 wherein the feature extraction module is configured to receive input signals associated to the user from a third sensor not associated to either the first airflow of the user or the second airflow of the user.

15. The system of claim 4 or 5 further comprising a headset, the headset comprising:

the first sensor positioned within at least the first airflow of the user;

the second sensor positioned within at least the second airflow of the user; and

a shielding member adapted to shield the first sensor from the second airflow, the shielding member adapted to provide an air gap between the shielding member and a face of the user while the headset is in use by the user.

16. The system of claim 15 wherein the headset further comprises a sensor assembly comprising a nasal sensor assembly and an oral sensor assembly, the nasal sensor assembly including the first sensor, the oral sensor assembly including the second sensor.

17. The system of claim 16 wherein the nasal sensor assembly and/or the oral sensor assembly include(s) respective Venturi structures shaped to case a constriction of a flow path within the nasal sensor assembly and/or the oral sensor assembly.

18. The system of any one of claims 15 to 17, the headset further comprising the third sensor located substantially outside of the first airflow of the user and the second airflow of the user.

19. A method of estimating speech airflow comprising:

receiving input signals associated to a user from at least a first sensor and a second sensor; and

determining an estimated shape and/or rate of airflow from at least part of the input signals.

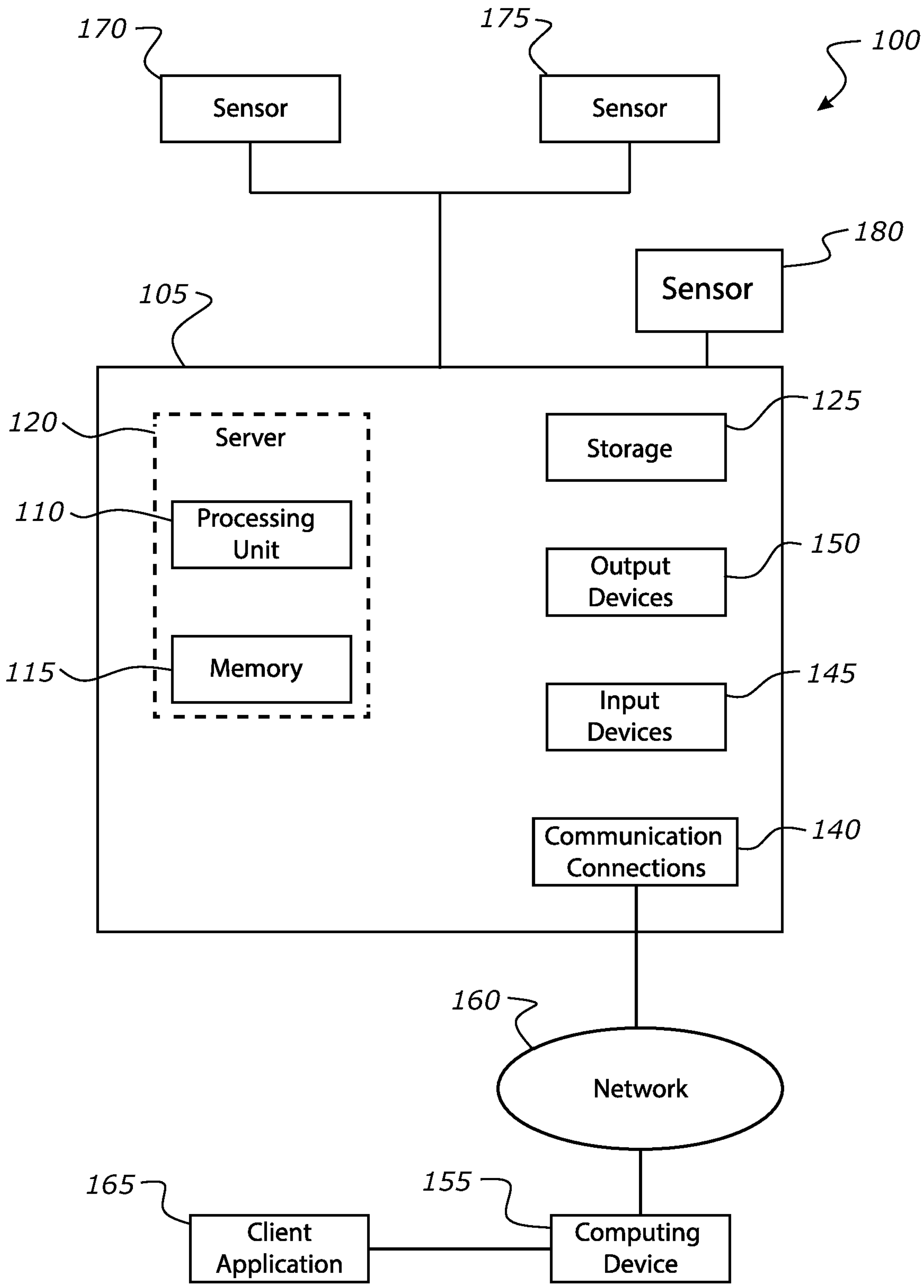
170
175
100
Sensor
Sensor
180
Sensor
105
120
Server
125
Storage
110
Processing Unit
150
Output Devices
115
Memory
145
Input Devices
140
Communication Connections
160
Network
165
155
Client Application
Computing Device

FIGURE 1

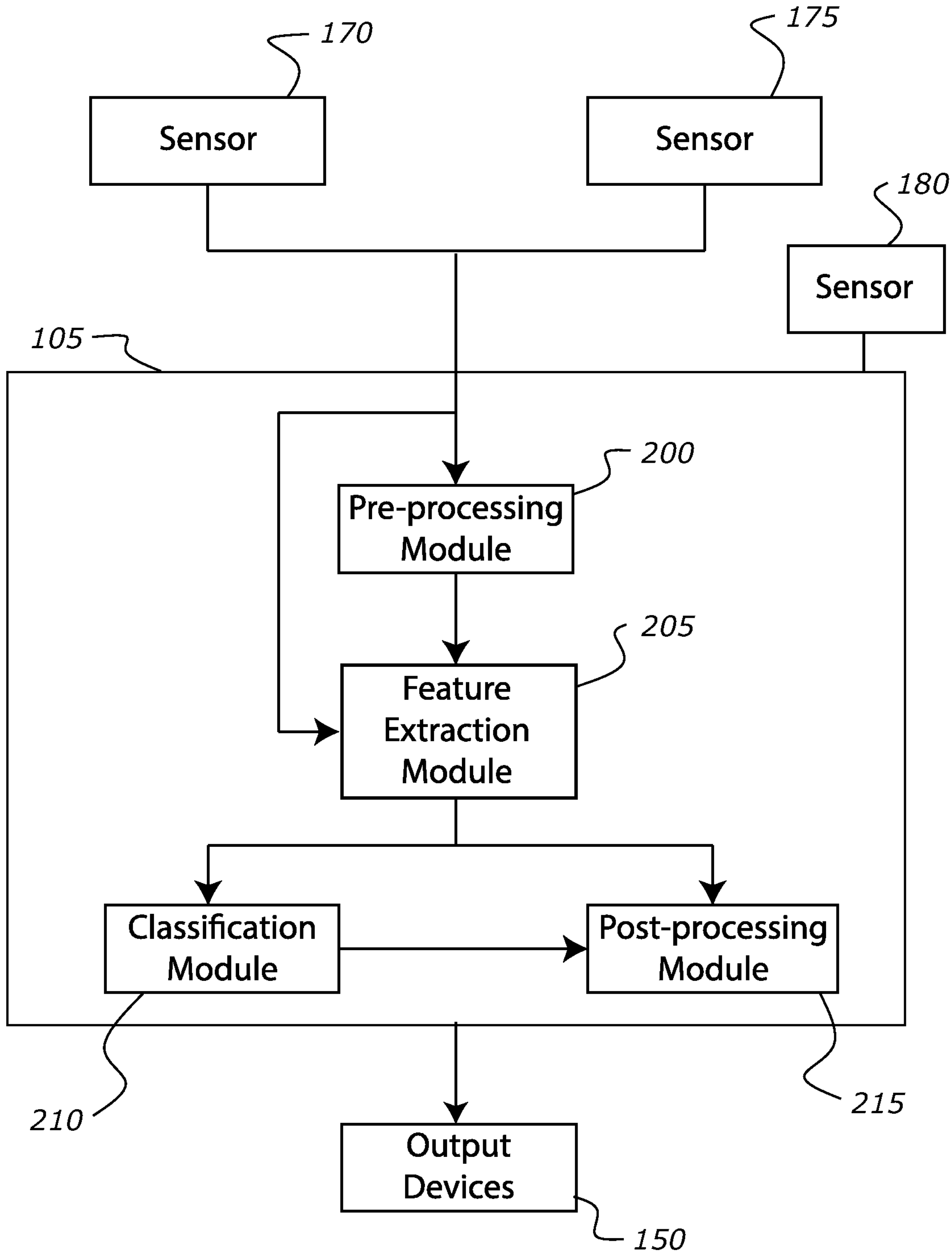
170
175
Sensor
Sensor
180
Sensor
105
200
Pre-processing Module
205
Feature Extraction Module
Classification Module
Post-processing Module
210
215
Output Devices
150

FIGURE 2

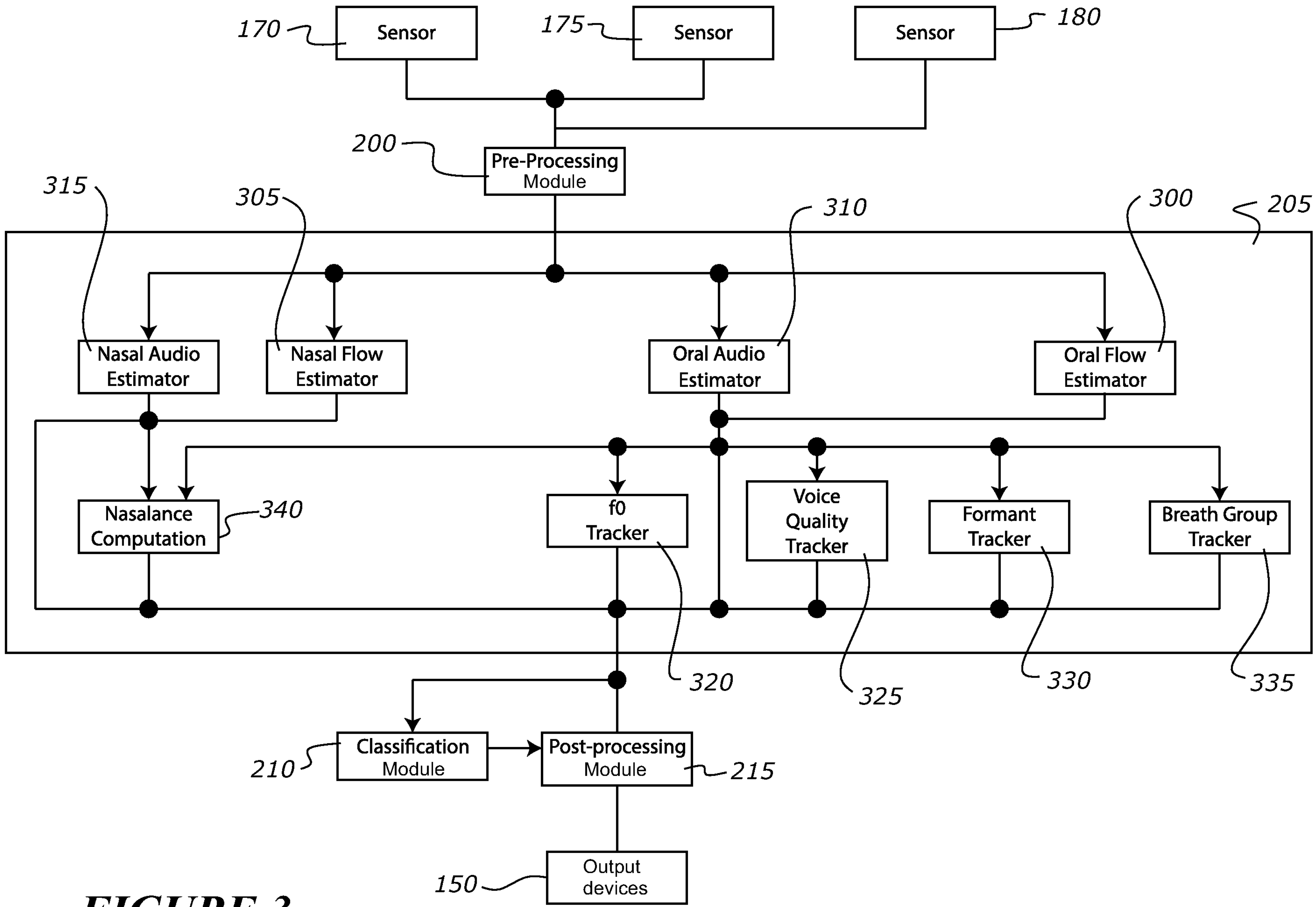
170
Sensor
175
Sensor
Sensor
180
200
Pre-Processing
Module
315
305
310
300
205
Nasal Audio
Estimator
Nasal Flow
Estimator
Oral Audio
Estimator
Oral Flow
Estimator
Nasalance
Computation
340
f0
Tracker
Voice
Quality
Tracker
Formant
Tracker
Breath Group
Tracker
320
325
330
335
Classification
Module
210
Post-processing
Module
215
150
Output
devices

FIGURE 3

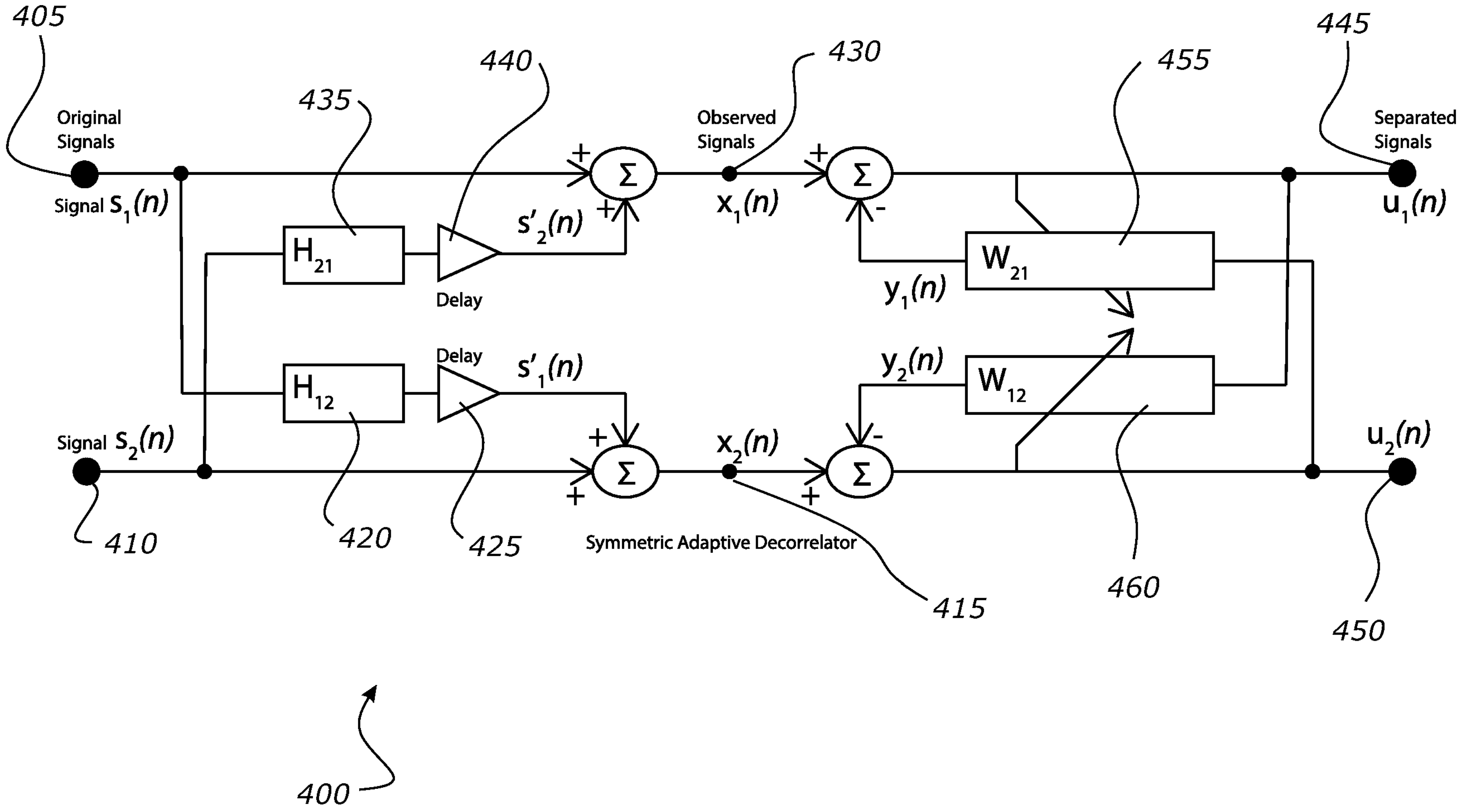
405
Original Signals
Signal s1(n)
435
440
H21
Delay
s'2(n)
+
Σ
Observed Signals
430
x1(n)
445
455
+
Σ
-
W21
y1(n)
Separated Signals
u1(n)
Signal s2(n)
H12
Delay
s'1(n)
+
Σ
+
x2(n)
y2(n)
W12
-
Σ
+
u2(n)
410
420
425
Symmetric Adaptive Decorrelator
415
460
450
400

FIGURE 4

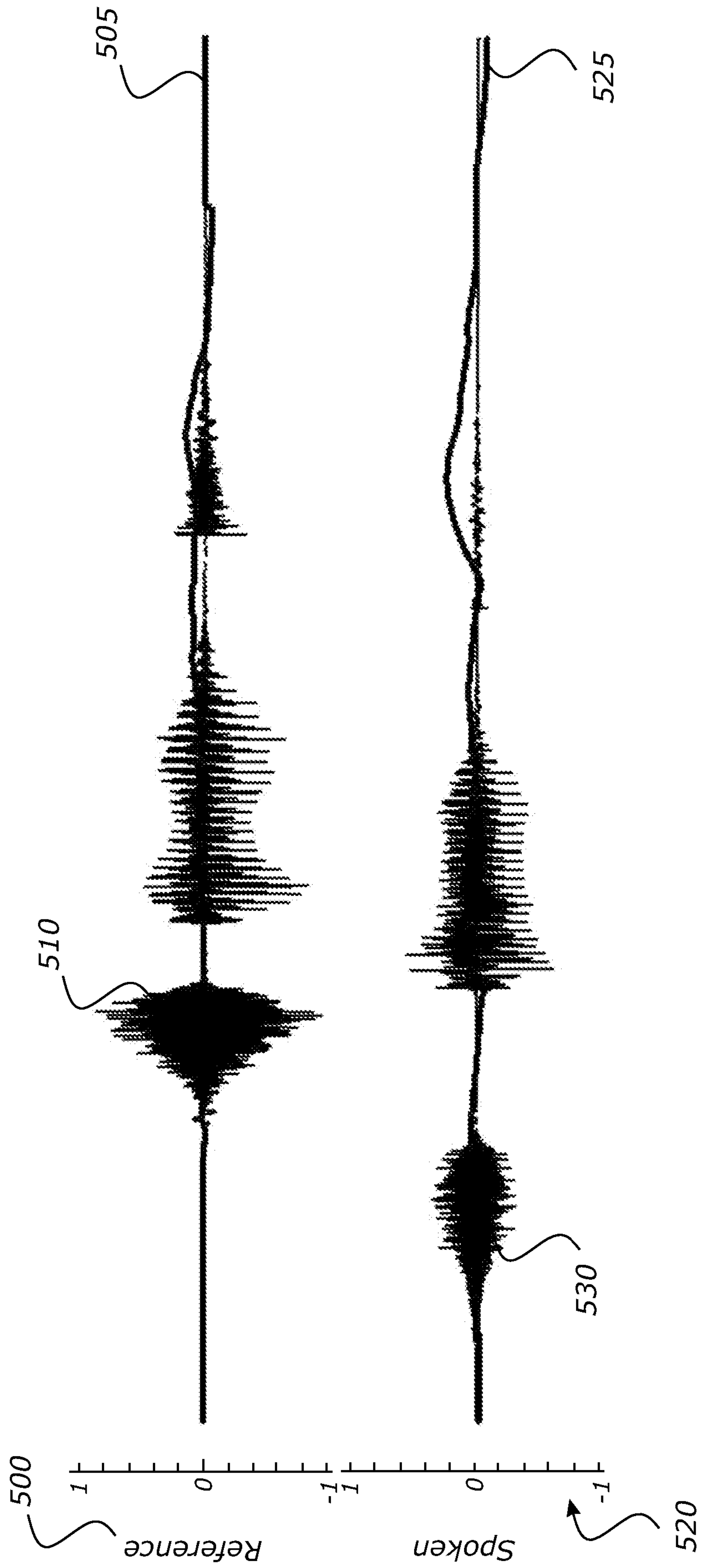

FIGURE 5

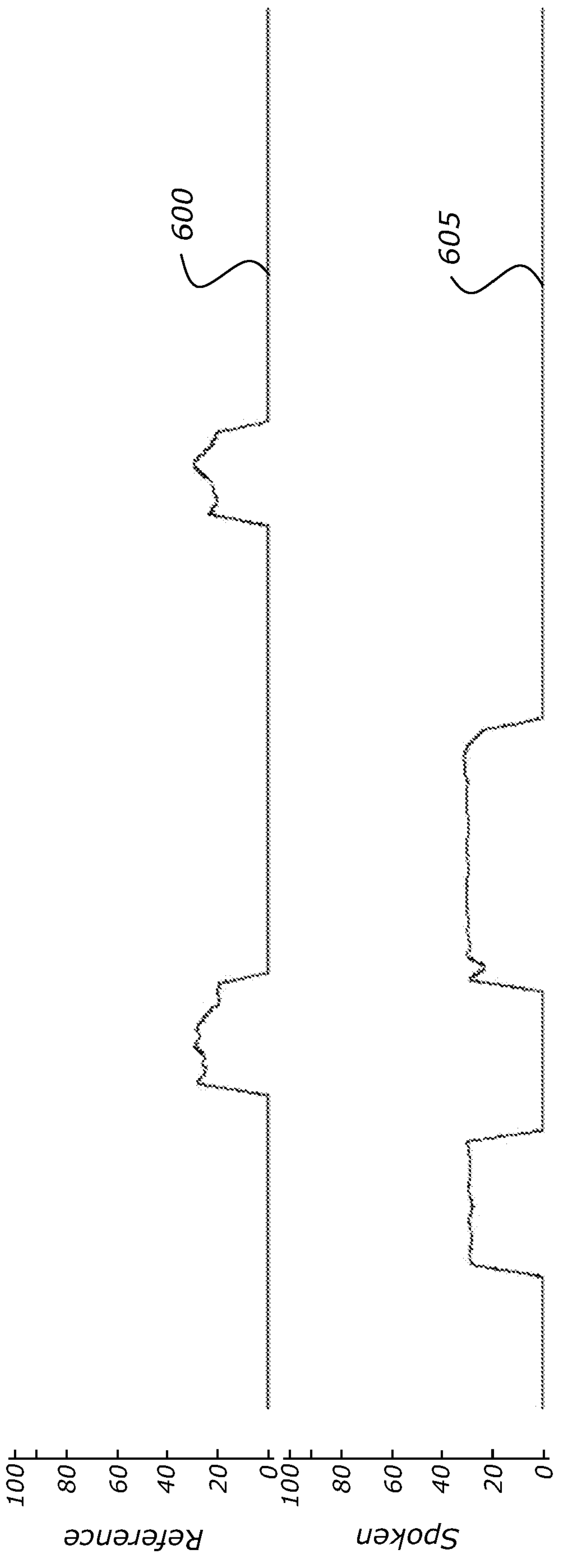

FIGURE 6

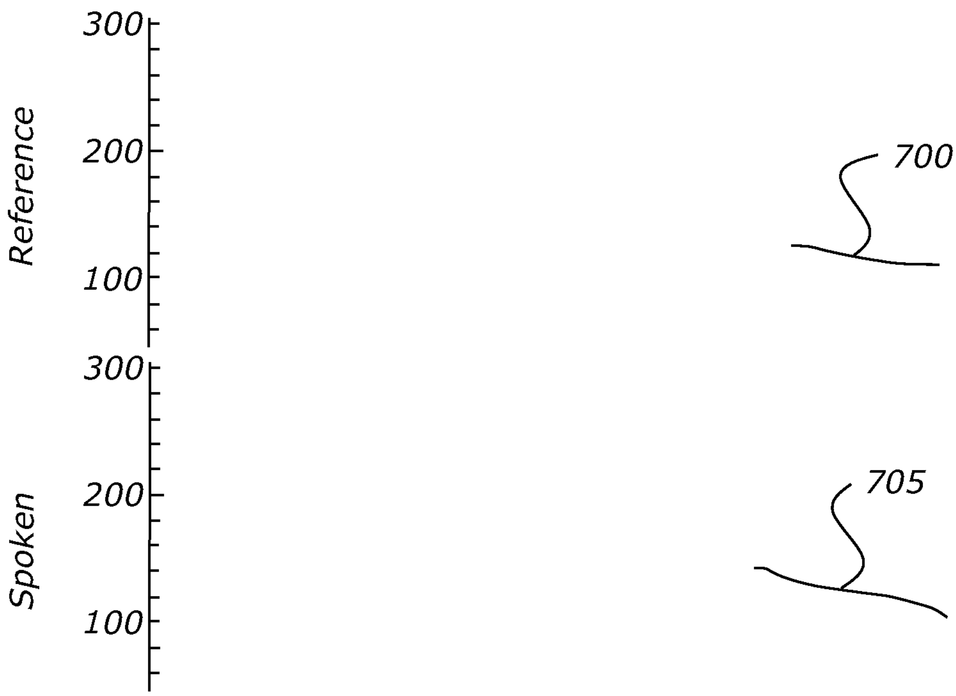

FIGURE 7

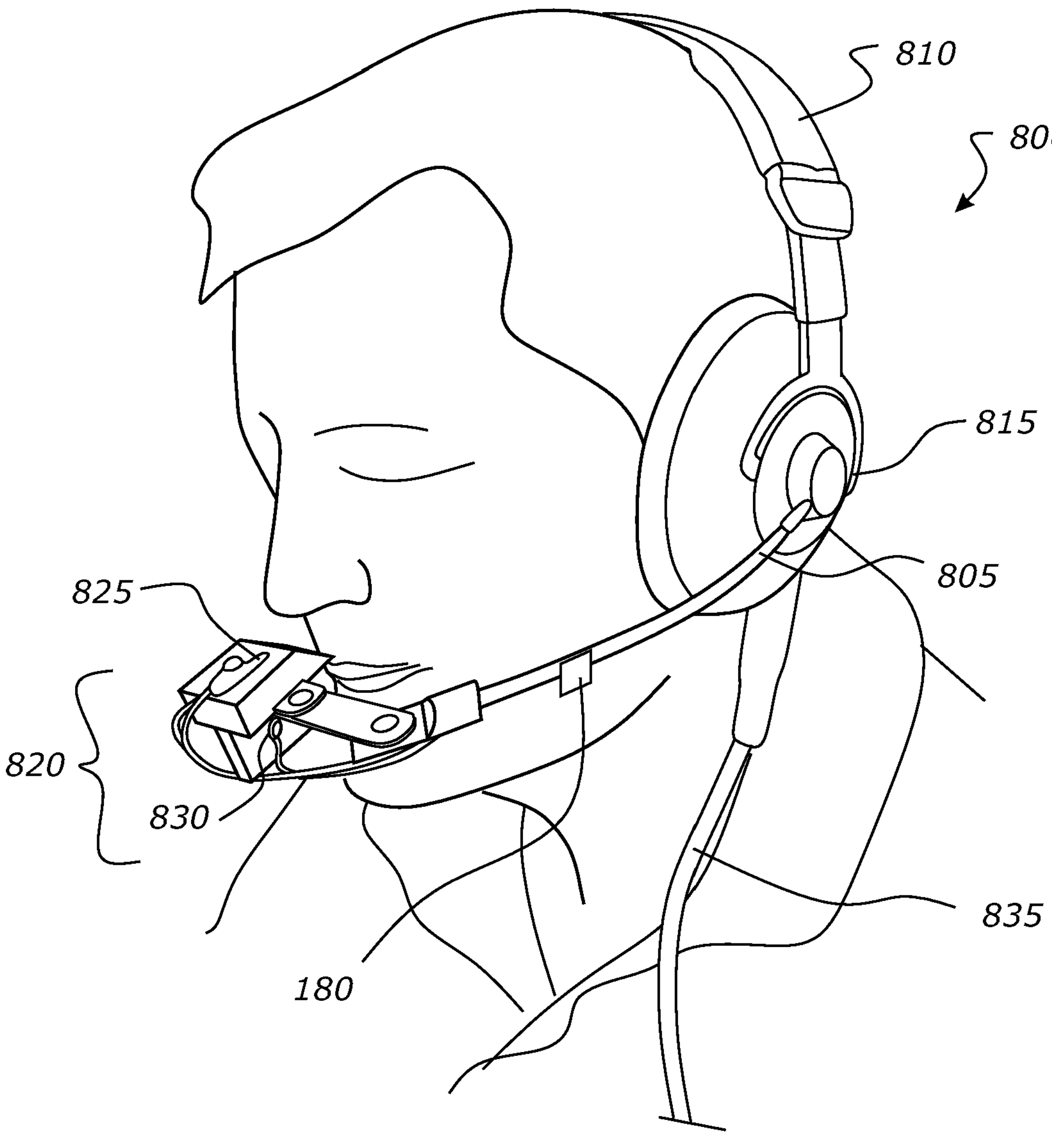

FIGURE 8

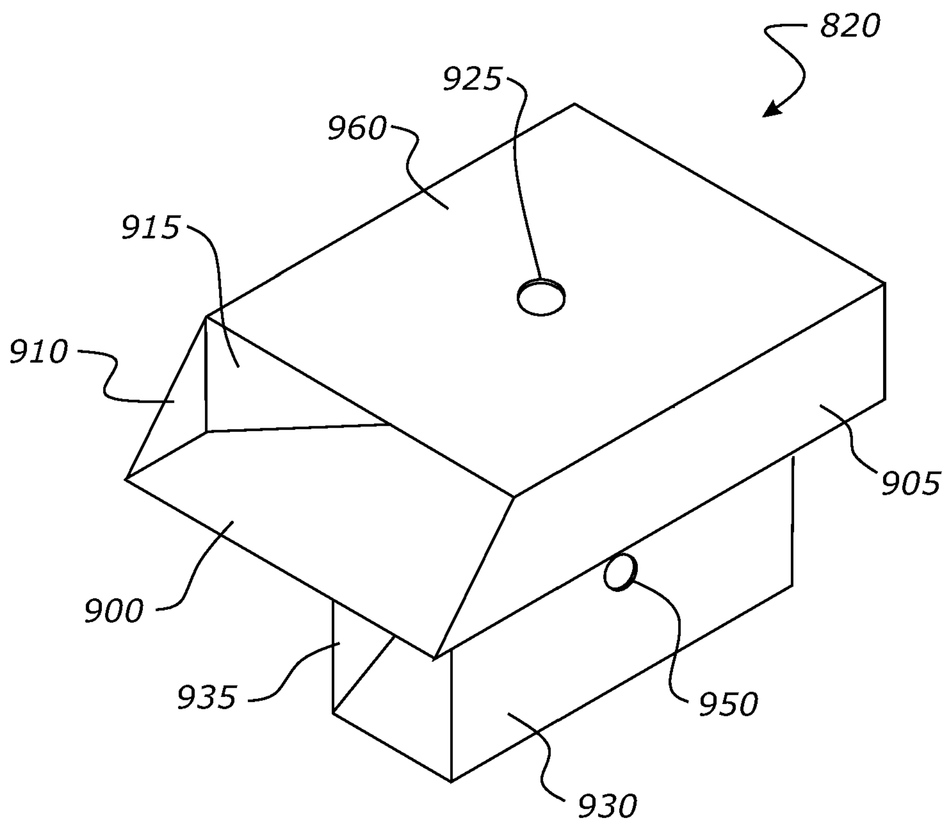

FIGURE 9

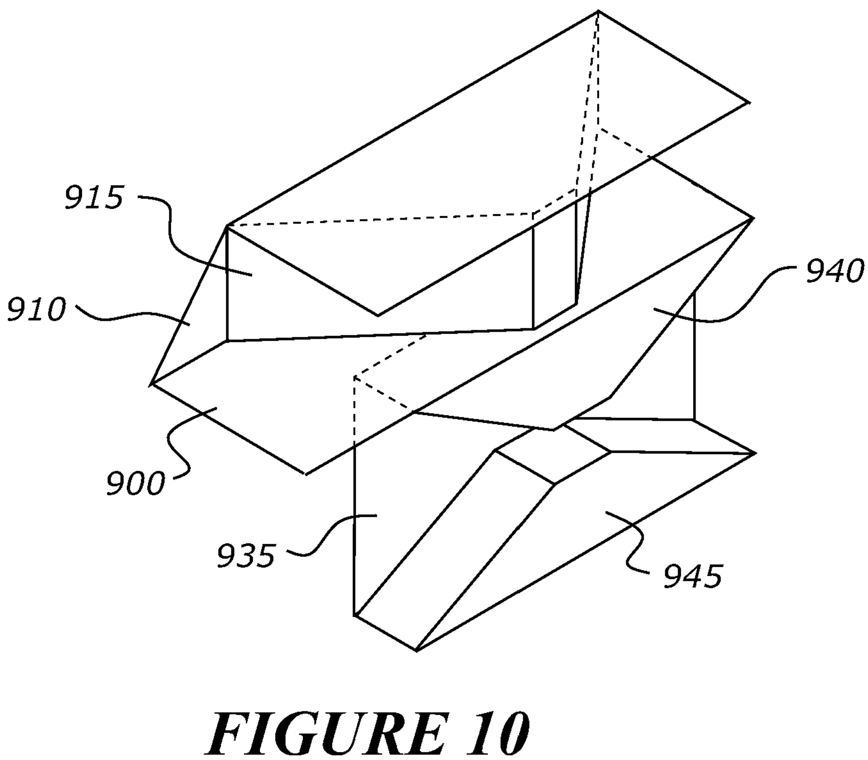

FIGURE 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/NZ2017/050104

A. CLASSIFICATION OF SUBJECT MATTER

A61B 5/087 (2006.01)

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)

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)

WPI, EPODOC: A61B5/--, G10L25/--, G10L21/--, G10L19/-- and search terms (speech, air, flow, sensor, extraction, voice, nasal, oral and like terms).

Applicant/Inventor name searched in internal databases provided by IP Australia.

Applicant/Inventor name searched in Espacenet.

Google Patents: Search words 'Speech Airflow Measure' and like terms.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

[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 November 2017	15 November 2017

Name and mailing address of the ISA/AU	**Authorised officer**
AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA Email address: pct@ipaustralia.gov.au	Timothy Williams AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No. +61262832067

INTERNATIONAL SEARCH REPORT

International application No.
PCT/NZ2017/050104

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 4519399 A (HORI) 28 May 1985 Abstract, Figures 2-9C, Column 2 Line 22-54, Column 3 Line 18-Column 5 Line 30	1-19
X	US 6850882 B1 (ROTHENBURG) 01 February 2005 Abstract, Figure 1, Column 3 Line 48-Column 8 Line 43, Column 10 Line 27-40	1-19
P,A	WO 2016/128627 A1 (AEROPHONOSCOPE) 18 August 2016 Whole Document	1-19
A	US 2016/0081611 A1 (HAMPTON et al.) 24 March 2016 Whole Document	1-19
A	US 2013/0090927 A1 (QUATIERI et al.) 11 April 2013 Whole Document	1-19

INTERNATIONAL SEARCH REPORT Information on patent family members	International application No. **PCT/NZ2017/050104**

This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document/s Cited in Search Report		**Patent Family Member/s**	
Publication Number	**Publication Date**	**Publication Number**	**Publication Date**
US 4519399 A	28 May 1985	US 4519399 A	28 May 1985
		DE 3345067 A1	20 Jun 1984
		FR 2537429 A1	15 Jun 1984
		FR 2537429 B1	25 Nov 1988
		JP S59107399 A	21 Jun 1984
US 6850882 B1	01 February 2005	US 6850882 B1	01 Feb 2005
WO 2016/128627 A1	18 August 2016	WO 2016128627 A1	18 Aug 2016
US 2016/0081611 A1	24 March 2016	US 2016081611 A1	24 Mar 2016
		US 2017112430 A1	27 Apr 2017
US 2013/0090927 A1	11 April 2013	US 2013090927 A1	11 Apr 2013
		US 9763617 B2	19 Sep 2017

End of Annex

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.