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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 December 2003 (24.12.2003)

PCT

(10) International Publication Number
WO 03/107327 A1

(51) International Patent Classification[7]: G10L 17/00, 15/26

(21) International Application Number: PCT/IB03/02345

(22) International Filing Date: 27 May 2003 (27.05.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
02077372.7 17 June 2002 (17.06.2002) EP

(71) Applicant *(for all designated States except US)*: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **VIGNOLI, Fabio** [IT/NL]; c/o Prof . Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: **GROENENDAAL, Antonius, W., M.**; Philips Intellectual Property & Standards, Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

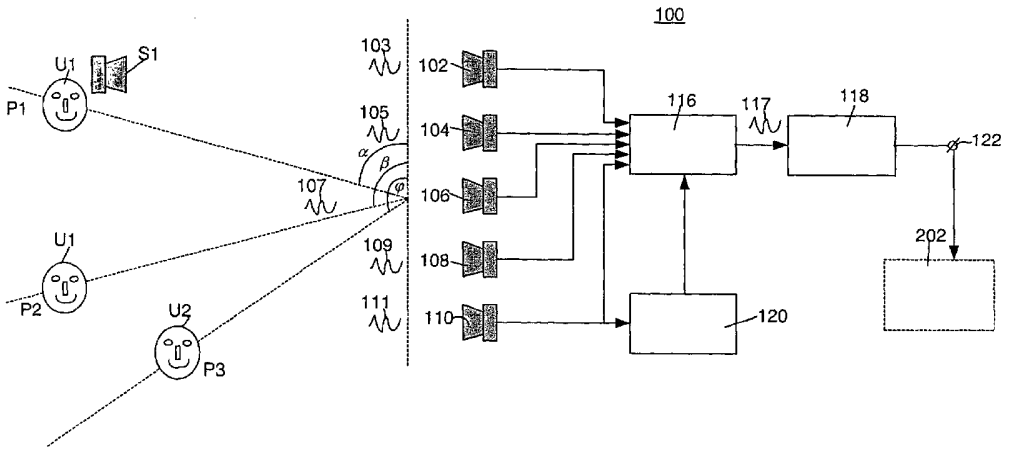
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

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

(54) **Title:** CONTROLLING AN APPARATUS BASED ON SPEECH



(57) **Abstract:** The speech control unit (100) comprises: a microphone array, comprising multiple microphones (102, 104, 106, 108 and 110) for receiving respective audio signals (103, 105, 107, 109 and 111); a beam forming module (116) for extracting a clean, i.e. speech, signal (117) of a user (U1), from the audio signals; a speaker recognition system (120) for recognition of the user (U1) based on a particular audio signal (111) and being arranged to control the beam forming module, on basis of the recognition; and a speech recognition unit (118) for creating an instruction for the apparatus (200) based on recognized speech items of the speech signal (117). As a consequence, the speech control unit (100) is more selective for those parts of the audio signals for speech recognition which correspond to speech items spoken by the user (U1).

WO 03/107327 A1

Controlling an apparatus based on speech

The invention relates to a speech control unit for controlling an apparatus on basis of speech, comprising:

- a microphone array, comprising multiple microphones for receiving respective audio signals;
- a beam forming module for extracting a speech signal of a user, from the audio signals as received by the microphones, by means of enhancing first components of the audio signals which represent an utterance originating from a first orientation of the user relative to the microphone array; and
- a speech recognition unit for creating an instruction for the apparatus based on recognized speech items of the speech signal.

The invention further relates to an apparatus comprising:

- such a speech control unit for controlling the apparatus on basis of speech; and
- processing means for execution of the instruction being created by the speech control unit.

The invention further relates to a method of controlling an apparatus on basis of speech, comprising:

- receiving respective audio signals by means of a microphone array, comprising multiple microphones;
- extracting a speech signal of a user, from the audio signals as received by the microphones, by means of enhancing first components of the audio signals which represent an utterance originating from a first orientation of the user relative to the microphone array; and
- creating an instruction for the apparatus based on recognized speech items of the speech signal.

Natural spoken language is a preferred means for human-to-human communication. Because of recent advances in automatic speech recognition, natural spoken language is emerging as an effective means for human-to-machine communication. The user is being liberated from manipulating a keyboard and mouse, which requires great hand/eye

coordination. This hands-free advantage of human to machine communication through speech recognition is particularly desired in situations where the user must be free to use his/her eyes and hands, and to move about unencumbered while talking. However the user is still encumbered in present systems by hand-held, body-worn, or tethered microphone equipment, e.g. headset microphone, which captures audio signals and provides input to the speech recognition unit. This is because most speech recognition units work best with a close-talking microphone input, e.g. with the user and microphone in close proximity. When they are deployed in "real-world" environments, the performance of known speech recognition units typically degrades. The degradation is particularly severe when the user is far from the microphone. Room reverberation and interfering noise contribute to the degraded performance.

In general it is uncomfortable to wear the headset microphone on a head for any extended period of time, while the hand microphone can limit freedom of the user as it occupies the user's hands, and there has been a demand for a speech input scheme that allows more freedom to the user. A microphone array in combination with a beam forming module appears to be a good approach that can resolve the conventionally encountered inconvenience described above. The microphone array is a set of microphones which are arranged at different positions. The multiple audio signals received by the respective microphones of the array are provided to the beam forming module. The beam forming module has to be calibrated, i.e. an orientation or position of a particular sound source relative to the microphone array has to be estimated. The particular sound source might be the source in the environment of the microphone array which generates sound having parameters corresponding to predetermined parameters, e.g. comprising predetermined frequencies matching with human voice. However, often the calibration is based on the loudest sound, i.e. the particular sound source generates the loudest sound. For example, a beam forming module can be calibrated on basis of the user who is speaking loudly, compared to other users in the same environment. A sound source direction or position can be estimated from time differences among signals from different microphones, using a delay sum array method or a method based on the cross-correlation function as disclosed in: "Knowing Who to Listen to in Speech Recognition: Visually Guided Beamforming", by U. Bub, et al. ICASSP'95, pp. 848-851, 1995. A parametric method estimating the sound source position (or direction) is disclosed in S. V. Pillai: "Array Signal Processing", Springer-Verlag, New York, 1989.

After being calibrated, i.e. the current orientation being estimated, the beam forming module is arranged to enhance sound originating from a direction corresponding to

the current direction and to reduce noise, by synthetic processing of outputs of these microphones. It is assumed that the output of the beam forming module is a clean signal that is appropriate to be provided to a speech recognition unit resulting in a robust speech recognition. This means that the components of the audio signals are processed such that the speech items of the user can be extracted.

An embodiment of a system comprising a microphone array, a beam forming module and a speech recognition unit is known from European Patent Application EP 0795851 A2. The Application discloses that a sound source position or direction estimation and a speech recognition can be achieved with the system. The disadvantage of this system is that it does not work appropriate in a multi user situation. Suppose that the system has been calibrated for a first position of the user. Then the user starts moving. The system should be re-calibrated first to be able to recognize speech correctly. The system requires audio signals, i.e. the user has to speak something, as input for the calibration. However, if in between another user starts speaking, then the re-calibration will not provide the right result: the system will get tuned to the other user.

It is an object of the invention to provide a speech control unit of the kind described in the opening paragraph which is arranged to recognize speech of a user who is moving in an environment in which other users might speak too.

The object of the invention is achieved in that the speech control unit comprises a speaker recognition system for recognition of the user based on a particular audio signal and being arranged to control the beam forming module, on basis of the recognition, in order to enhance second components of the audio signals which represent a subsequent utterance originating from a second orientation of the user relative to the microphone array. In this case the speaker recognition system is used to discriminate between audio signals related to utterances of different users. The speech control unit is arranged to re-calibrate if it receives sound originating from a recognized user, from a different orientation. Preferably this recognized user is the user who initiated an attention span (see also Fig. 3) of the apparatus to be controlled. There will be no re-calibration on basis of a non-recognized user. As a consequence, speech items spoken by a non-recognized user will be discarded.

In an embodiment of the speech control unit according to the invention, the speaker recognition system is arranged to recognize another user based on the particular audio signal and is arranged to control the beam forming module, on basis of this recognition, in order to enhance third components of the audio signals which represent another utterance originating from a third orientation of the other user relative to the microphone array. This embodiment of the speech control unit is arranged to re-calibrate on basis of another recognized user. Besides, following one particular user, this embodiment is arranged to calibrate on basis of sound from a selected set of users. That means that only authorized users, i.e. those who have authorization to control the apparatus, are recognized as such and hence only speech items from them will be accepted for the creation of instructions for the apparatus.

In an embodiment of the speech control unit according to the invention, a first one of the microphones of the microphone array is arranged to provide the particular audio signal to the speaker recognition system. In other words, the particular audio signal which is used for speaker recognition corresponds to one of the audio signals as received by the microphones of the microphone array. The advantage is that no additional microphone is required.

In an embodiment of the speech control unit according to the invention, the beam forming module is arranged to determine a first position of the user relative to the microphone array. Besides orientation, also a distance between the user and the microphone array is determined. The position is calculated on basis of the orientation and distance. An advantage of this embodiment according to the invention is that the speech control unit is arranged to discriminate between sounds originating from users who are located in front of each other.

It is a further object of the invention to provide an apparatus of the kind described in the opening paragraph which is arranged to be controlled by a user who is moving in an environment in which other users might speak too.

This object of the invention is achieved that the apparatus comprises the speech control unit as claimed in claim 1.

An embodiment of the apparatus according to the invention is arranged to show that the user has been recognized. Preferably the apparatus is also arranged to show which user has been recognized, in the case of multiple users is providing speech items to the apparatus.

An embodiment of the apparatus according to the invention which is arranged to show that the user has been recognized, comprises audio generating means for generating an audio signal representing the user. By generating an audio signal comprising a representation of the name of the user, e.g. "Hello Jack" it is clear for the user that the apparatus is ready to receive speech items from the user. This concept is also known as auditory greeting.

An embodiment of the apparatus according to the invention which is arranged to show that the user has been recognized, comprises a display device for displaying a visual representation of the user. By displaying a personalized icon or an image of the user it is clear for the user that the apparatus is ready to receive speech items from the user. In other words, the apparatus is in an active state of classifying and/or recognizing speech items.

An embodiment of the apparatus according to the invention which is arranged to show that the user has been recognized, is developed to show a set of controllable parameters of the apparatus on basis of a preference profile of the user. Many apparatus have numerous controllable parameters. However not all of these controllable parameters are of interest for each user of the apparatus. Besides that, each of the users has his own preferred default values. Hence, a user has a so-called preference profile. It is advantageously to show the default values of the controllable parameters which are of interest to the user, e.g. the user who initiated the attention span.

It is a further object of the invention to provide a method of the kind described in the opening paragraph which enables to recognize speech of a user who is moving in an environment in which other users might speak too.

This object of the invention is achieved that the method is characterized in comprising recognition of the user based on a particular audio signal and controlling the extraction of the speech signal of the user, on basis of the recognition, in order to enhance second components of the audio signals which represent a subsequent utterance originating from a second orientation of the user relative to the microphone array.
Modifications of speech control unit and variations thereof may correspond to modifications and variations thereof of the apparatus described and of the method described.

These and other aspects of the speech control unit, of method and of the apparatus according to the invention will become apparent from and will be elucidated with

respect to the implementations and embodiments described hereinafter and with reference to the accompanying drawings, wherein:

Fig. 1 schematically shows an embodiment of the speech control unit according to the invention;

Fig. 2 schematically shows an embodiment of the apparatus according to the invention; and

Fig. 3 schematically shows the creation of an instruction on basis of a number of audio signals.

Same reference numerals are used to denote similar parts throughout the figures.

Fig. 1 schematically shows an embodiment of the speech control unit 100 according to the invention. The speech control unit 100 is arranged to provide instructions to the processing unit 202 of the apparatus 200. These instructions are provided at the output connector 122 of the speech control unit 100, which comprises:

- a microphone array, comprising multiple microphones 102, 104, 106, 108 and 110 for receiving respective audio signals 103, 105, 107, 109 and 111;
- a beam forming module 116 for extracting a clean, i.e. speech, signal 117 of a user U1, from the audio signals 103, 105, 107, 109 and 111 as received by the microphones 102, 104, 106, 108 and 110;
- a speaker recognition system 120 for recognition of the user U1 based on a particular audio signal 111 and being arranged to control the beam forming module, on basis of the recognition; and
- a speech recognition unit 118 for creating an instruction for the apparatus 200 based on recognized speech items of the speech signal 117.

The working of the speech control unit 100 is as follows. It is assumed that initially the speech control unit 100 is calibrated on basis of utterances of user U1 being at position P1. The result is that the beam forming module 116 of the speech control unit 100 is "tuned" to sound originating from directions which substantially match direction α. Sound from directions which differ from direction α with more than a predetermined threshold, is disregarded for speech recognition. E.g. speech of user U2, being located at position P2 with a direction φ relative to the microphone array is neglected. Preferably, the speech control unit 100 is sensitive to sound with voice characteristics, i.e. speech, and is insensitive to

others sounds. For instance the sound of the music as generated by the speaker S1, which is located in the vicinity of user U1 is filtered out by the beam forming module 116.

Suppose that user U1 has moved to position P1, corresponding to an orientation β relative to the microphone array. Without re-calibration of the speech control unit 100, or more particular the beam forming module 116, the recognition of speech items probably would fail. However the speech control unit 100 will get calibrated again when user U1 starts speaking. The utterances of user U1 are recognized as being generated by U1 and used for the re-calibration. If another user, e.g. user U2, starts speaking then his/her utterances are recognized as not being from user U1 and skipped for the re-calibration. As a consequence the speech control unit 100 is arranged to stay "tuned" to user U1 while he/she is moving. Speech signals of this user U1 are extracted from the audio signals 103, 105, 107, 109 and 111 and are basis for speech recognition. Other sounds are not taken into account for the control of the apparatus.

Above it is explained that the speech control unit 100 is arranged to "follow" one specific user U1. This user might be the user who initiated the attention span. Optionally, the speech control unit 100 is arranged to get subsequently tuned to a number of users, being part of a "known" group of people, i.e. which will be recognized as being members of that group. In that case the speech control unit 100 is designed to accept only speech items of these persons and not from other persons.

In Fig. 1 is depicted that the microphone 110 is connected to both the speaker recognition system 120 and the beam forming module 116. This is optional, that means that an additional microphone could have been used.. The components 116-120 of the speech control unit 100 and the processing unit 202 of the apparatus 200 may be implemented using one processor. Normally, both functions are performed under control of a software program product. During execution, normally the software program product is loaded into a memory, like a RAM, and executed from there. The program may be loaded from a background memory, like a ROM, hard disk, or magnetically and/or optical storage, or may be loaded via a network like Internet. Optionally an application specific integrated circuit provides the disclosed functionality.

Fig. 2 schematically shows an embodiment of the apparatus 200 according to the invention. The apparatus 200 optionally comprises audio generating means 206 for generating an audio signal representing the user U1. By generating an audio signal comprising a representation of the name of the user U1, e.g. "Hello Jack" it is clear for the user U1 that the apparatus is ready to receive speech items from the user U1. In other words,

the apparatus is in an active state of recognizing speech items. The generating means 206 comprises a memory device for storage of a sampled audio signal, a sound generator and a speaker. Optionally, the apparatus also comprises a display device 204 for displaying a visual representation of the user U1. By displaying a personalized icon or an image of the user it is clear for the user that the apparatus is ready to receive speech items from the user U1.

The speech control unit 100 according to the invention is preferably used in a multi-function consumer electronics system, like a TV, set top box, VCR, or DVD player, game box, or similar device. But it may also be a consumer electronic product for domestic use such as a washing or kitchen machine, any kind of office equipment like a copying machine, a printer, various forms of computer work stations etc, electronic products for use in the medical sector or any other kind of professional use as well as a more complex electronic information system. Besides that, it may be a product specially designed to be used in vehicles or other means of transport, e.g. a car navigation system. Whereas, the word "multifunction electronic system" as used in the context of the invention may comprise a multiplicity of electronic products for domestic or professional use as well as more complex information systems, the number of individual functions to be controlled by the method would normally be limited to a reasonable level, typically in the range from 2 to 100 different functions. For a typical consumer electronic product like a TV or audio system, where only a more limited number of functions need to be controlled, e.g. 5 to 20 functions, examples of such functions may include volume control including muting, tone control, channel selection and switching from inactive or stand-by condition to active condition and vice versa, which could be initiated, by control commands such as "louder", "softer", "mute", "bass" "treble" "change channel", "on", "off", "stand-by" etcetera.

In the description it is assumed that the speech control unit 100 is located in the apparatus 200 being controlled. It will be appreciated that this is not required and that the control method according to the invention is also possible where several devices or apparatus are connected via a network (local or wide area), and the speech control unit 100 is located in a different device then the device or apparatus being controlled.

Fig. 3 schematically shows the creation of an instruction 318 on basis of a number of audio signals 103, 105, 107, 109 and 111 as received by the microphones 102, 104, 106, 108 and 110. From the audio signals the speech items 304-308 are extracted. The speech items 304-308 are recognized and voice commands 312-316 are assigned to these speech items 304-308. The voice commands 312-316 are “Bello”, “Channel” and “Next”,

respectively. An instruction "Increase_Frequency_Band", which is interpretable for the processing unit 202 is created based on these voice commands 312-316.

To avoid that conversations or utterances not intended for controlling the apparatus are recognized and executed, the speech control unit 100 optionally requires the user to activate the speech control unit 100 resulting in a time span, or also called attention span during which the speech control unit 100 is active. Such an activation may be performed via voice, for instance by the user speaking a keyword, like "TV". By using an anthropomorphic character a barrier for interaction is removed: it is more natural to address the character instead of the product, e.g. by saying "Bello" to a dog-like character. Moreover, a product can make effective use of one object with several appearances, chosen as a result of several state elements. For instance, a basic appearance like a sleeping animal can be used to show that the speech control unit 100 is not yet active. A second group of appearances can be used when the speech control unit 100 is active, e.g. awake appearances of the animal. The progress of the attention span can then, for instance, be expressed, by the angle of the ears: fully raised at the beginning of the attention span, fully down at the end. The similar appearances can also express whether or not an utterance was understood: an "understanding look" versus a "puzzled look". Also audible feedback can be combined, like a "glad" bark if a speech item has been recognized. A user can quickly grasp the feedback on all such system elements by looking at the one appearance which represents all these elements. E.g. raised ears and an "understanding look", or lowered ears and a "puzzled look".

Once a user has started an attention span the apparatus, i.e. the speech control unit 100 is in a state of accepting further speech items. These speech items 304-308 will be recognized and associated with voice commands 312-316. A number of voice commands 312-316 together will be combined to one instruction 318 for the apparatus. E.g. a first speech item is associated with "Bello", resulting in a wake-up of the television. A second speech item is associated with the word "channel" and a third speech item is associated with the word "next". The result is that the television will switch, i.e. get tuned to a next broadcasting channel. If another user starts talking during the attention span of the television just initiated by the first user, then his/her utterances will be neglected.

It should be noted that the above-mentioned embodiments illustrate rather than limit the invention and that those skilled in the art will be able to design alternative embodiments without departing from the scope of the appended claims. In the claims, any reference signs placed between parentheses shall not be constructed as limiting the claim. The word 'comprising' does not exclude the presence of elements or steps not listed in a

claim. The word "a" or "an" preceding an element does not exclude the presence of a plurality of such elements. The invention can be implemented by means of hardware comprising several distinct elements and by means of a suitable programmed computer. In the unit claims enumerating several means, several of these means can be embodied by one and the same item of hardware.

CLAIMS:

1. A speech control unit (100) for controlling an apparatus (200) on basis of speech, comprising:

- a microphone array, comprising multiple microphones (102, 104, 106, 108 and 110) for receiving respective audio signals (103, 105, 107, 109 and 111);

- a beam forming module (116) for extracting a speech signal (117) of a user (U1), from the audio signals as received by the microphones, by means of enhancing first components of the audio signals which represent an utterance originating from a first orientation (α) of the user (U1) relative to the microphone array; and

- a speech recognition unit (118) for creating an instruction for the apparatus (200) based on recognized speech items of the speech signal, characterized in comprising a speaker recognition system (120) for recognition of the user based on a particular audio signal (111) and being arranged to control the beam forming module (116), on basis of the recognition, in order to enhance second components of the audio signals which represent a subsequent utterance originating from a second orientation (β) of the user (U1) relative to the microphone array.

2. A speech control unit (100) as claimed in claim 1, characterized in that the speaker recognition system (120) is arranged to recognize another user (U2) based on the particular audio signal (111) and is arranged to control the beam forming module (116), on basis of this recognition, in order to enhance third components of the audio signals which represent another utterance originating from a third orientation (φ) of the other user relative to the microphone array.

3. A speech control unit (100) as claimed in claim 1, characterized in that a first one of the microphones (110) of the microphone array is arranged to provide the particular audio signal (111) to the speaker recognition system (120).

4. A speech control unit (100) as claimed in claim 1, characterized in that the beam forming module (116) is arranged to determine a first position of the user relative to the microphone array.

5. An apparatus (200) comprising:

- a speech control unit (100) for controlling the apparatus (200) on basis of speech as claimed in claim 1; and

- processing means (202) for execution of the instruction being created by the speech control unit (100).

6. An apparatus (200) as claimed in claim 5, characterized in being arranged to show that the user has been recognized.

7. An apparatus (200) as claimed in claim 6, characterized in comprising audio generating means (206) for generating an audio signal representing the user.

8. An apparatus (200) as claimed in claim 6, characterized in comprising a display device (204) for displaying a visual representation of the user (U1).

9. An apparatus (200) as claimed in claim 6, characterized in being arranged to show a set of controllable parameters of the apparatus (200) on basis of a preference profile of the user (U1).

10. A consumer electronics system comprising the apparatus (200) as claimed in claim 5.

11. A method of controlling an apparatus (200) on basis of speech, comprising:

- receiving respective audio signals by means of a microphone array, comprising multiple microphones;

- extracting a speech signal of a user, from the audio signals as received by the microphones, by means of enhancing first components of the audio signals which represent an utterance originating from a first orientation of the user relative to the microphone array; and

- creating an instruction for the apparatus (200) based on recognized speech items of the speech signal, characterized in comprising recognition of the user based on a particular audio signal and controlling the extraction of the speech signal of the user, on basis of the recognition, in order to enhance second components of the audio signals which represent a subsequent utterance originating from a second orientation of the user relative to the microphone array.

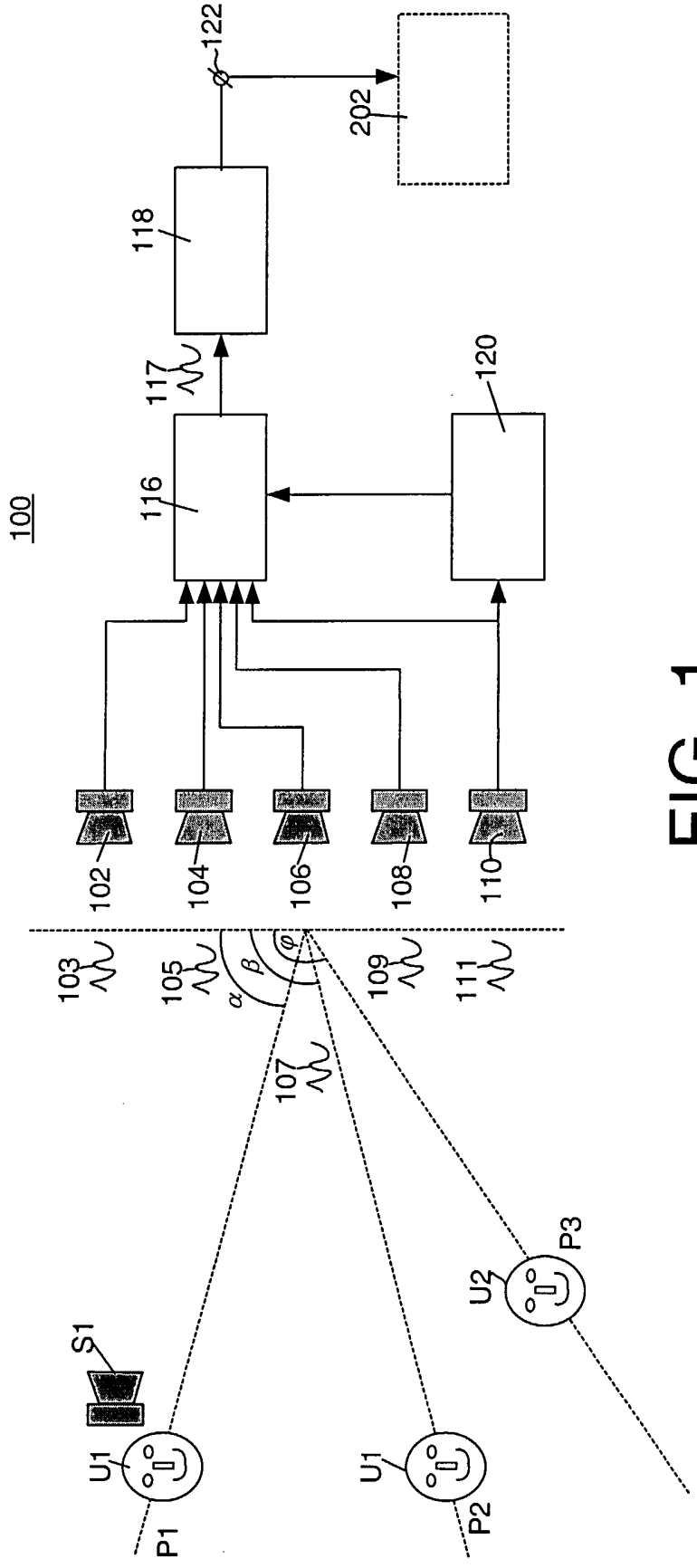

100
102
104
106
108
110
103
105
109
111
107
α
β
φ
116
117
118
120
122
202
S1
U1
U1
U2
P1
P2
P3

FIG. 1

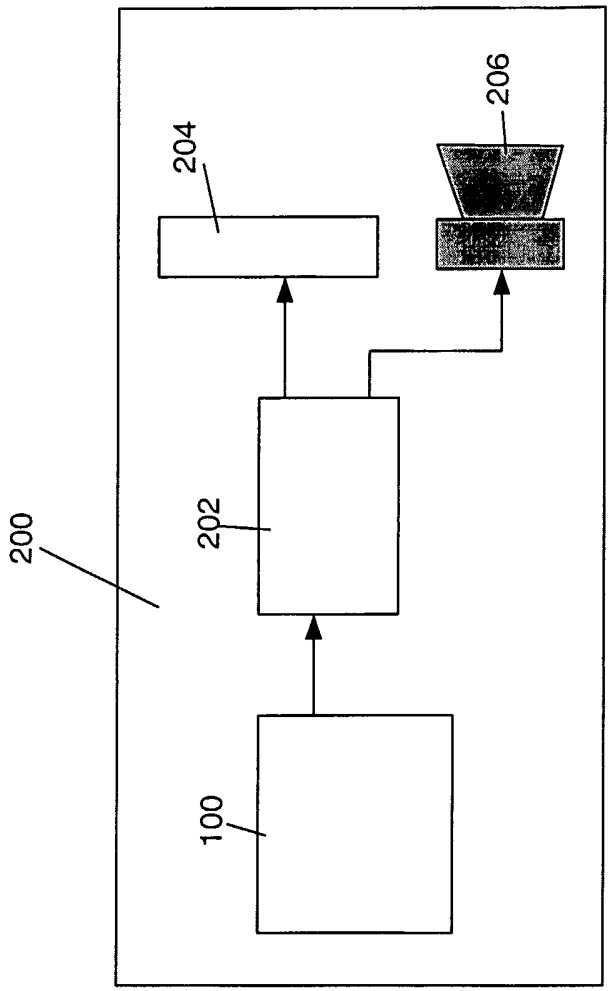


FIG. 2

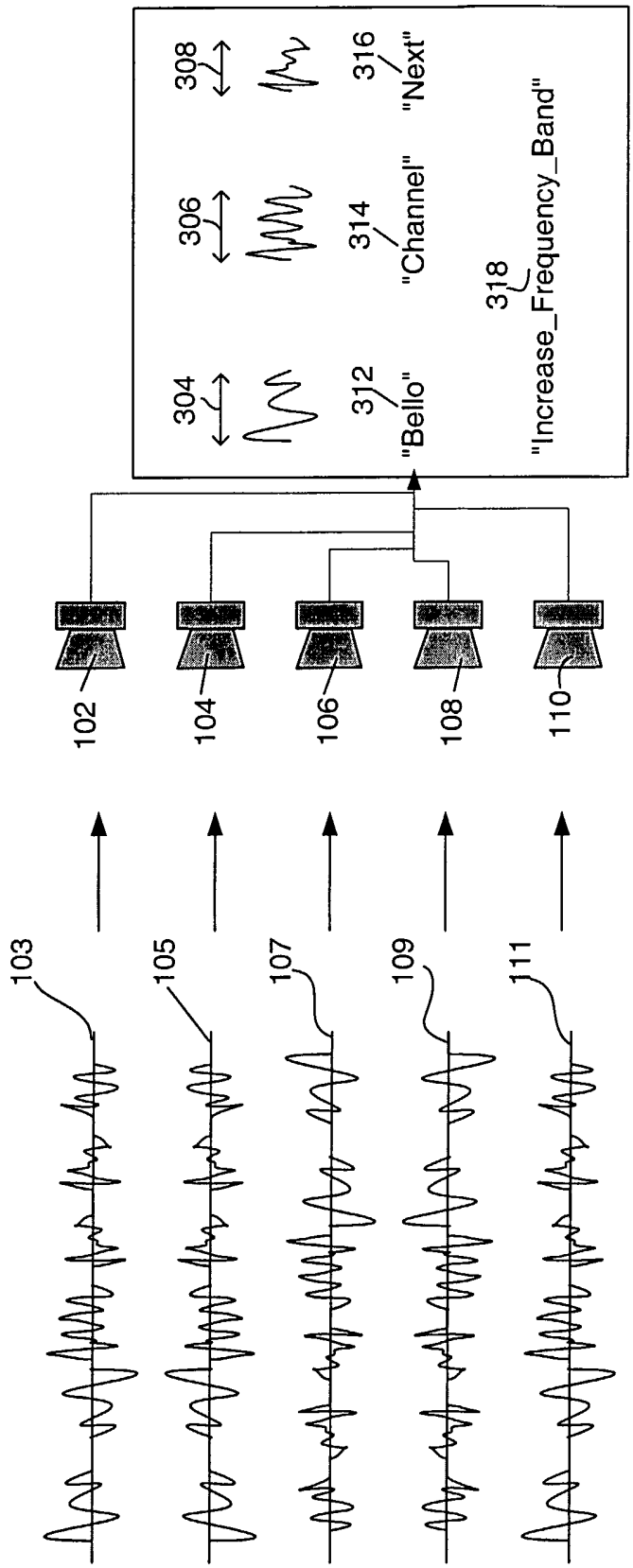

304
306
308
312
"Bello"
314
"Channel"
316
"Next"
318
"Increase_Frequency_Band"
102
104
106
108
110
103
105
107
109
111

FIG. 3

INTERNATIONAL SEARCH REPORT

Int ial Application No
PCT/IB 03/02345

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L17/00 G10L15/26

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)
IPC 7 G10L

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 practical, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LLEIDA E ET AL: "Robust continuous speech recognition system based on a microphone array" ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 1998. PROCEEDINGS OF THE 1998 IEEE INTERNATIONAL CONFERENCE ON SEATTLE, WA, USA 12-15 MAY 1998, NEW YORK, NY, USA,IEEE, US, 12 May 1998 (1998-05-12), pages 241-244, XP010279154 ISBN: 0-7803-4428-6	1-5,10, 11
Y	Chapters Introduction, 2 and 4	6,8

-/--

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

[ ] Patent family members are listed in 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 document 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
5 September 2003	22/09/2003

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, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Bourdier, R

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	TRIVEDI M ET AL: "Intelligent environments and active camera networks" SYSTEMS, MAN, AND CYBERNETICS, 2000 IEEE INTERNATIONAL CONFERENCE ON NASHVILLE, TN, USA 8-11 OCT. 2000, PISCATAWAY, NJ, USA,IEEE, US, 8 October 2000 (2000-10-08), pages 804-809, XP010524751 ISBN: 0-7803-6583-6 page 808, left-hand column, line 1 -page 809, left-hand column, line 12 ---	6,8
A	TAKAYUKI NAGAI ET AL: "Estimation of source location based on 2-D music and its application to speech recognition in cars" INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, ICASSP 2001, vol. 5, 7 - 11 May 2001, pages 3041-3044, XP002253587 Salt Lake City page 3041, column 1, line 1 -page 3041, column 2, line 2 -----	1,2,5, 10,11